

*From mechanistic to statistical
modeling, then mechanistic again:*
Systems-biology approaches uncover key
host-cell determinants of susceptibility to
Coxsackievirus B3

A Dissertation

*Presented to the faculty of the School of Engineering
and Applied Science in partial fulfillment of the
requirements for the degree of*

Doctor of Philosophy

By Andrew J. Sweatt

December 2023

Department of Biomedical Engineering
University of Virginia

APPROVAL SHEET

This

is submitted in partial fulfillment of the requirements
for the degree of

Author:

Advisor:

Advisor:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

A handwritten signature in black ink, appearing to read "J. L. West". The signature is fluid and cursive, with the first letter of each name being capitalized and prominent.

Jennifer L. West, School of Engineering and Applied Science

Table of Contents and Figures

Acknowledgements.....	7
Abstract	8
Chapter 1: Introduction.....	9
1.1 Background.....	10
1.2 CVB3: its life cycle, genetics, and biochemistry	11
1.3 The immune response to CVB3 and implications in myocarditis	12
1.3.1 The host-cell immune response	12
1.3.2 The innate and adaptive immune responses	13
1.4 Systems virology	14
1.5 Overview of dissertation	16
1.6 References	17
 Chapter 2: Modeling the complete kinetics of coxsackievirus B3 reveals human determinants of host-cell feedback[#]	 24
2.1 Abstract.....	25
2.2 Introduction.....	25
2.3 Results	26
2.3.1 Modular draft of a complete kinetic model for the CVB3 life cycle	26
Figure 2.1. A modular encoding of the CVB3 life cycle elaborated with antiviral responses and viral antagonism of host-cell processes.	28
Figure 2.2. Control genes for dsRNA induction and representative immunoblots of the cycloheximide experiment quantified in Figure 2.1C.....	29
2.3.2 Explanatory modeling of CVB3 tropism requires careful extracellular bookkeeping	30
Figure 2.3. Stoichiometric estimation and simulation of CVB3 particles, cargo, and surface receptors.....	33
Figure 2.4. Delivery controls, immunoblot quantification, V5 immunoblot image, and evaluation of receptor–ratio thresholds.....	35
2.3.3 A two-phase encoding of membrane replication predicts quantifiable positive–negative strands and replicative intermediates.....	36
Figure 2.5. Explosive genome replication requires intracellular membranes and hides negative-strand templates from quantitation.	39
Figure 2.6. Polysome size requirements, strand-specific tagged qPCR validation, and population-level simulations of viral RNA strand dynamics with complete kinetics.....	40
2.3.4 Encapsidation must coordinate the kinetics of enteroviral protein synthesis, recruitment, and self-assembly	42
Figure 2.7. A Goldilocks zone for enteroviral encapsidation.	45
Figure 2.8. Stoichiometric balance of CVB3 protein classes and population-level simulations of viral-protein and eIF4G cleavage dynamics.....	47

2.3.5 Loss of type I interferon signaling coincides with degradation of MAVS during CVB3 infection	48
Figure 2.9. CVB3 partially dismantles antiviral signaling and the dsRNA transducer MAVS.	51
Figure 2.10. Controls for interferon-stimulated genes and time-dependent antagonism of mature virion formation by supplemental interferon.	53
2.3.6 A cleavage-resistant MAVS mutant shows enhanced antiviral activity upon delayed stimulation with paracrine interferon	54
Figure 2.11. MAVS is a sensitive locus for CVB3 susceptibility in the host-cell network.	56
Figure 2.12. Poisson noise of the plaque assay and generality of the timed-interferon effect to Ala¹⁴⁸Ala²⁷¹-mutant MAVS.	58
2.3.7 Enteroviral proteinase cleavage of MAVS is redirected by a prevalent human polymorphism.....	59
Figure 2.13. MAVS is a sensitive locus for CVB3 susceptibility in the human population.	61
Figure 2.14. Sensitivity analysis of the MAVS regulation model.	63
2.4 Discussion	64
2.5 STAR Methods	67
2.6 Appendix	85
Figure SN1. Annotated dynamics of the MAVS filamentation model.	87
2.7 References.....	88
 Chapter 3: Simulating Coxsackievirus B3 infection with an accessible computational model of its complete kinetics	96
3.1 Abstract.....	97
3.2 Before you begin.....	97
3.2.1 Background and implementation	97
3.2.2 Install the graphical user interface (GUI)	99
Figure 3.1. MacOS folder setup for MATLAB Runtime.	103
3.3 Key resources table.....	103
3.4 Step-by-step method details.....	104
3.4.1 Basic operations.....	104
Figure 3.2. Demonstration of a single simulation and sequential simulations.	105
Figure 3.3. Turning off the antiviral response and plotting a different species.	108
Figure 3.4. Adding exogenous IFN stimulation.	109
3.4.2 Population mode.....	110
Figure 3.5. A simulation in population mode.	111
3.4.3 Sensitivity analysis.....	112
Figure 3.6. Performing a single-parameter sensitivity analysis.	113
3.5 Expected outcomes.....	114
Figure 3.7. Model GUI outputs for the selected species during a viral life cycle.	115
Figure 3.8. Model GUI outputs for selected species identify a threshold viral polymerase rate for productive infection.	115

Figure 3.9. Model GUI outputs comparing single cell mode to population mode at a low MOI.	117
3.6 Quantification and statistical analysis.....	118
Figure 3.10. Model GUI outputs comparing serial simulations with different run counts.....	119
3.7 Limitations.....	120
3.8 Troubleshooting.....	122
3.9 Tables	125
Table 3.1: General Tab Input Options	125
Table 3.2. Options Tab Input Options.....	126
Table 3.3. Advanced Options Tab Input Options.....	127
3.10 References.....	132
Chapter 4: Proteome-wide copy-number estimation from transcriptomics#	134
4.1 Foreword.....	135
4.2 Abstract.....	136
4.3 Introduction.....	136
4.4 Results	138
4.4.1 Deriving three gene-specific biological classes of mRNA-protein relationships	138
Figure 4.1. Meta-assembly and inference of conditional mRNA-to-protein relationships for 4366 human genes.	141
Figure 4.2. Calibration for the proteomic meta-assembly and model selection examples.	142
Figure 4.3. Pinferna model selection is consistent with known biological mechanisms and mRNA-to-protein relationships.....	143
Figure 4.4. Extended characterization of M and HL mRNA-to-protein relationship classes.....	144
4.4.2 Pinferna predictions in cell lines and tissues relative to competing alternatives	145
Figure 4.5. Pinferna outperforms empirical guesses and competing methods for absolute protein abundance estimation.....	147
Figure 4.6. Robustness of Pinferna predictions.....	148
4.4.3 Application to in silico modeling.....	149
Figure 4.7. Simulating degrees of human cardio-susceptibility to coxsackievirus B3 (CVB3) infection based on inferred abundance differences in CVB3 receptors....	151
Figure 4.8. Calibrated protein inferences of CD55 and CXADR yield disease-related predictions of coxsackievirus B3 (CVB3) susceptibility.....	152
4.4.4 Application to molecular subtyping	154
Figure 4.9. Consensus re-clustering of 796 breast cancer cases from The Cancer Genome Atlas (TCGA) by protein inference.....	155
Figure 4.10. Inferred proteomics reassigns luminal A/B transcriptomic subtypes of breast cancer.....	156
4.5 Discussion	158
4.6 Materials and Methods	160
4.7 Appendix Supplementary Methods.....	167

4.8 References.....	170
Chapter 5: Discussion.....	176
5.1 Summary, limitations, broader impacts, and future directions.....	177
Figure 5.1. The complete kinetics model can be quickly adapted to other viruses.	180
5.2 Unpublished, but never forgotten	183
Figure 5.2. Proof of concept for AAV2/9 transduction of V5-MAVS-P2A-3xFLAG-luciferase (luc) into Mavs^{-/-} mice.	185
Figure 5.3. Inverse temperature dependence of human rhinovirus 2 (HRV2).....	187
Figure 5.4. Implications of IRF3 Arg96 or Gln96 in innate immune signaling.....	189
5.3 Concluding remarks	192
5.4 References.....	193

Acknowledgements

Wow, I'm finally here. My dissertation. I couldn't have done this without the support of some key people:

First, to my mentor and advisor, Dr. Kevin Janes. It's no secret that my grad school prospects were limited... Kevin's lab, or a place (that will remain unnamed) that I really didn't want to go. It's also no secret that Kevin took a gamble when he accepted me into his lab—I didn't have research experience, and I came from an unknown university. But he saw drive in me. Thank you, Kevin, for giving me the opportunity to show what I can do. What I thought being a scientist was and what I have learned it is are two different things. You have truly become a role model of what it means to be an excellent scientist and mentor.

Second, to the Janes lab through the years. Each person created such a welcoming environment and made me excited to go to lab every day. Y'all have pushed me and grown me as a scientist, and for that I am deeply thankful. To Taylor, who has been with me since day 1... literally, since we started on the same day—our friendship has been a great support during my PhD. To Wisam—I've really enjoyed our friendship and all the recipe swapping, Marvel movie watching, and board games. To Cameron—you are one of my science role models. You really love it, and you're really good at it. Thank you for joining me on the virus side of the lab. To Róza—you are one of the most encouraging people I know. Thank you for welcoming me into your home and family. Thank you for constantly encouraging me in my projects, for providing great advice, and being one of the most positive people I know.

Third, to my church family at the Blue Ridge Church. A PhD is hard. Like, really hard. Not to mention a good chunk of mine was during the pandemic. Thank you, church family, for lending listening ears, even when you didn't know what you were listening to. Thank you for constant encouragement. Thank you for deep friendships. To Richard and Keegan Sawyer—y'all may not know it, but little interactions and times that we've hung out have meant a lot to me. Thank you for welcoming me into your family. To the Bridge ministry—thank you for giving me the opportunity to serve you. To be honest, I was exhausted most Fridays, but you provided me with so much energy and fun. I think the ministry fed me more than I could feed it. I loved every minute of it. To Phil and Beth Bruns—y'all have been incredibly supportive. I can't think of two people more on the side of young professionals than you two. Thank you for giving me the opportunity to serve the church in the song ministry. Thank you for encouragement. Thank you for advice. Thank you for your humble hearts to learn, serve, teach, lead and advise. I want to be you when I grow up. To Stephen "Sbebo" and Liza Whetzel—y'all's friendship got me through a lot. We shared a love of food and cooking, of board games, of outdoor adventures, of Bean, of Jesus. I could always count on being filled when I was with y'all. To Finn and Leah Smyth—thank you for keeping my bad-decision undergrad days alive by staying up into the wee hours of the morning carried away in deep conversation about our fears, hopes, and dreams for ourselves and the church. To Will and Kelly-Jayne Portillo—thank you for being solid friends who can so quickly and so willingly switch to deep conversations about God. Thank you for modeling what it means to never stop seeking relationship.

Fourth, to Stephen and Erin Lees—I've truly relied on our times together. There has never been a time I didn't look forward to being together. During my PhD, I've had more fun and adventure with y'all in the short time I've known you than I had combined before knowing you. These times have much needed relaxation. I have similarly been inspired by your love for God and people. When I think of servants, you come to mind. Your love inspires me. To Stephen, specifically—my friendship soul mate. I could write a dissertation about what your friendship has meant to me. But I don't want to write another dissertation. One was enough. So I'll just say, thank you. I look forward to a lifetime of friendship.

Fifth, to friends afar and abroad. To Jed, Ashley, Caleb and Eli Veatch—when I think of a servant family who looks out for others, you are the blueprint. In college, you took me in. You encouraged, fed me, provided a place to stay at times, invited me into your family entirely. Though talks are rare, I'll never forget a thing. I get carried away looking at photos old and new of the boys. To Kelly and Kennedy Vance, and Ryan and Charity Cagle—y'all are pretty much the reason I had fun in college. And our friendship goes way deep than that. Thank you for always being on my side and being willing to listen.

Finally, to my family. I had no bigger cheerleaders than my parents. Even though they didn't really ever know what I was talking about, they always asked and wanted to know what the latest was in research. They were always interested in all things life. For years of encouraging me and letting me learn how to balance having fun and working, thank you. Many might not think it, but that is a skill, and an important one for a PhD. Thank you, and I love you.

Abstract

Viruses are important agents of disease. Though we know much about the life cycles, genetics, and biochemistry of many viruses, the link between infection and disease remains elusive. There have been studies into the importance of host genetics and immunity in determining outcomes to viral infection, but it has not been considered whether small biochemical changes in host-cell species can alter infection dynamics and resulting disease. These sorts of studies require a broad view of the biological system since viral proteins alter the cell state to favor replication while cell-signaling feedback loops prevent or counteract the virus-induced cell state. My goal is to elucidate how population-level differences in host-cell species alter viral infection and resulting disease.

My work focused on coxsackievirus B3 (CVB3), a causative agent of cardiac inflammation (myocarditis). CVB3 is one of the best studied viruses, with a clearly defined life cycle, extensive information on each of its 11 proteins, and decades of research into its role in myocarditis. Thus, CVB3 is perhaps one the best model systems for taking a systems-biology approach to studying virus–host interactions.

I began by building a computational mechanistic model for the entire CVB3 life cycle with host-cell immune responses overlaid as negative feedbacks. Using the model, we uncovered a sensitivity to the timing of the type I interferon response that is dependent on host-cell resistance to cleavage of the key innate-immune signaling protein MAVS by viral proteinases. Looking further into MAVS, we identified a polymorphism at amino acid 93 that dictates susceptibility to cleavage. We show computationally and experimentally that the polymorphism is able to modulate the severity of CVB3 infection. Thus, MAVS is one host-cell species where a small population-level difference can have a big impact on disease.

I then asked how differences in the abundances of the CVB3 receptors DAF and CAR impact susceptibility to infection. Rather than sample over an artificial range of abundances, I obtained RNA-seq data from 1489 human heart samples. However, RNA-seq is not a good substitute for protein abundances. To obtain protein-level estimates, I developed statistical models able to predict protein abundance from mRNA abundance. I predicted the paired abundances of DAF and CAR. When used in the mechanistic model for CVB3, I identified individuals with varying degrees infection severity and cases where individuals were completely resistant to infection. The results demonstrate how intrinsic differences in protein levels can have large impacts on susceptibility to infection.

This dissertation thus identifies two key host-cell determinants with population-level variability that dictate susceptibility to viral infection. The finding was only possible using a systems-biology approach that blended mechanistic and statistical modeling.

Chapter 1:

Introduction

1.1 Background

Viruses are one of the most prevalent causes of disease and death worldwide. To date, 260+ viruses are known to infect humans (1) and cause diseases such as deadly hemorrhagic fevers (e.g., ebolavirus) (2), pneumonia (e.g., influenza) (3), hand-foot-and-mouth disease (e.g., enterovirus 71) (4), gastroenteritis (“stomach bug,” e.g., rotavirus) (5), and “the cold” (e.g., rhinovirus) (6). Though there are hundreds of viruses, Nobel-laureate virologist David Baltimore broadly classified them into seven groups based on whether their genome i) consists of DNA or RNA, ii) is single- or double-stranded, and iii) is positive- or negative-sense. Regardless of its class, a virus’s life cycle is generally the same: a virus particle (virion) docks to and enters the host cell; the genome escapes; translation results in viral proteins that perturb host-cell functions, replicate the virus, and package new virions; and new virions are released from the host to infect neighboring cells (7). For many viruses, the details of each step are well characterized. Yet, the link between viral infection and disease remains elusive (8). Take, for example, SARS-CoV-2, the virus responsible for COVID-19. We know the entry mechanism (9), the functions of most of the viral proteins (9), and how the genome has evolved throughout its existence in the human population (10). However, knowledge of the virus alone is not enough to explain the heterogeneity of COVID-19 presentation. The severity of COVID-19 is determined by host genetic factors (11, 12), the timing and level of cytokine release (13), and quality of the immune response (14). Disease, therefore, comprises direct viral damage, the host-cell response, and the immune system operating in a (dys)regulated fashion. While SARS-CoV-2 is currently a notorious example, every virus has complex interactions with hosts that together define disease.

Enteroviruses are a genus of non-enveloped, single-stranded, positive-sense RNA viruses responsible for a host of diseases (15). Details of their genetics, proteins and life cycles are found in most virology textbooks (e.g., (16)), and they have been studied for 70+ years (17). Despite this, hundreds of enterovirus serotypes persist and impact people differently for reasons that are unclear (18). There have been studies into the importance of host genetics in infection (19, 20), and the importance of immunity is appreciated. However, it has not been considered whether small biochemical changes in host-cell species can alter infection dynamics and resulting disease. My goal is to elucidate how differences in host-cell species alter enteroviral infection and resulting disease.

In this dissertation, I focus on a prototypical enterovirus, coxsackievirus B3 (CVB3). As a disease agent, CVB3 has been widely studied as a causative agent of viral myocarditis, or inflammation of the heart (21–24). Upon infection of cardiomyocytes, many proinflammatory pathways are initiated, resulting in expression of cytokines in heart tissue (25). Cytokines recruit immune cells to fight the virus and clear infected cells. Generally,

infections are resolved in the acute infection stage, but some progress to a chronic phase that can lead to heart failure (26, 27). At the cellular level, host pathologies are the result of virus–host interactions. During infection, cell-signaling feedback loops prevent or counteract the virus-induced cell state to fight the infection (28–31). The dozens of interactions are daunting but approachable using systems biology (7, 32).

In the remainder of this introduction, I will provide necessary background into CVB3 and provide details on host immunity to the virus. From there, I will briefly review how systems biology has been used in virology, then conclude with a broader picture of this dissertation.

1.2 CVB3: its life cycle, genetics, and biochemistry

The steps of the CVB3 life cycle are well defined and broadly encompass three stages: entry, replication, and encapsidation. Entry begins with CVB3 binding to decay-accelerating factor (DAF) on the surface of cardiomyocytes (33). DAF:CVB3 clustering promotes signaling through the kinases Abl and Fyn. Abl promotes translocation of DAF:CVB3 clusters to tight junctions (34), where CVB3 unbinds DAF and binds to coxsackie and adenovirus receptor (CAR) (34–36). CAR:CVB3 undergoes endocytosis via caveolae through signaling events triggered by Fyn (34). Once internalized, the single-stranded, positive-sense viral RNA (+vRNA) genome is released into the cytoplasm.

The +vRNA undergoes cap-independent translation into a single polypeptide by host ribosomes that recognize an internal ribosome entry site (37). The proteinases 2A (2A^{pro}) and 3C (3C^{pro}) autocatalytically cleave themselves from the polypeptide and process the rest of the polypeptide (16). The viral proteins 2B, 2C^{ATPase}, and 3A (and precursor 3AB) co-opt host-cell membranes to form viral replication organelles (VROs) that tether replication machinery and support replication (38–41). Replication is carried out by the RNA-dependent RNA polymerase 3D^{pol}. 3D^{pol} recognition of vRNA is mediated by viral protein 3B^{VPg} and the *cis*-active RNA element structure of the genome (42, 43). +vRNA is used as a template to synthesize negative-sense (–)vRNA, and vice versa. Imbalances in replication give rise to excess +vRNA (44) for packaging.

The final step is encapsidation of free +vRNA into new virions. The structural proteins VP0, VP1, and VP3 bind to form a protomer, and five protomers bind to form a pentamer, the base unit of the capsid (45). The VP3 subunit of a pentamer interacts with 2C^{ATPase} (46), recruiting pentamers to the VRO membranes where excess +vRNA is being synthesized. +vRNA is inserted into forming capsids by a mechanism not fully known (45). Ultimately, 12 pentamers bind to form a new virion, and cleavage of the VP0 subunits into VP2 and VP4 signal the end of the process (45). Accumulation of virions causes cell lysis

~8 hours following internalization.

In the midst of the viral life cycle, the 2A^{pro} and 3C^{pro} disrupt host-cell processes to create a proviral environment and inhibit the host-cell immune response. The proteinases cleave eukaryotic translation initiation factor 4G and PABP to halt host protein synthesis (47–49), which liberates ribosomes for translation of the CVB3 genome. Host-cell immune proteins are also cleaved (50–52), preventing signaling through type I interferons. Further, trafficking is disrupted, nucleocytoplasmic shuttling integrity is lost, and membrane structural integrity is lost (53–56). Altogether, the events alter the cell state to favor efficient viral replication and release.

1.3 The immune response to CVB3 and implications in myocarditis

Most of the time, thinking of the immune response conjures up antibodies and white blood cells such as macrophages and T cells. However, the first step of the pathogenic immune response is for the infected host cell to recognize and attempt to fight the pathogen by turning on many cell-signaling pathways. From there, the innate and adaptive arms of the immune system are recruited.

1.3.1 The host-cell immune response

At the nexus of the antiviral cell-signaling pathway is the mitochondrial antiviral signaling protein (MAVS). MAVS (also IPS-1, VISA, and Cardif) quickly became the subject of much study when four seminal papers were published within six weeks of one another in 2005 (57–60). All four papers showed the same conclusion—MAVS induces and is required for the type I interferon response. Since then, MAVS has been characterized extensively. At its 3' terminus, MAVS is docked to the mitochondria (59). At its 5' terminus, MAVS contains a caspase activation and recruitment domain (CARD) (57–59). MAVS is activated upon association with CARD-containing helicases in the RIG-I family (57, 58). The RIG-I family member melanoma differentiation-associated protein 5 (MDA5) is responsible for detecting picornaviruses (61) and associates with MAVS to drive MAVS polymerization on the mitochondrial surface (62). Once polymerized, various domains on MAVS become active. The TRAF2 and TRAF6 binding sites recruit their respective TRAF, which activate a proinflammatory response through NF- κ B activation (63, 64) and recruit TANK-binding kinase 1 (TBK1) (65, 66). TBK1 phosphorylates the pLxIS motif on MAVS, which recruits inactive interferon regulatory factor 3 (IRF3). IRF3—now in proximity to TBK1—is phosphorylated, homodimerizes, translocates to the nucleus, and transcribes type I interferons (IFN) (66). Type I IFNs signal in an autocrine–paracrine fashion through the JAK/STAT pathway to turn on interferon-stimulated genes (ISGs) (29). Hundreds of ISGs

switch the cell towards an antiviral state, but they generally perform a few main functions: shutting down translation to prevent viral protein synthesis, degradation of (viral and host) RNA, and preventing endocytic trafficking to stop entry of new virus particles (29). Other targets include MDA5 (67), establishing a positive feedback loop to maintain the antiviral state. The host-cell immune response through type I IFN signaling is a crucial step of the pathogenic immune response partly responsible for preventing mortality and promoting viral clearance.

1.3.2 The innate and adaptive immune responses

Type I IFNs and other cytokines recruit the “classical” immune systems arms: the innate and adaptive systems. Innate immune cells are nonspecific first responders to damage or pathogens and secrete cytokines that further the host-cell immune response and contribute to the cytokine milieu that regulates the recruitment and function of immune cells. The adaptive immune system consists (most generally) of T cells and B cells. Unlike innate immune cells, T and B cells “adapt” to the pathogen and thus have highly specific responses to clear the infection. Adaptive immune cells also provide long-term protection against pathogens through formation of memory T cells and antibody-producing B cells. Together, innate and adaptive immune cells provide a highly organized and robust response to clear the pathogen.

Macrophages and NK cells are the primary innate immune cells that mediate CVB3-induced myocarditis (26). In general, macrophages are professional phagocytes of apoptotic cells (68) and are major producers of nitric oxide (NO), which has an antiviral effect (69). At the same time, macrophages are partly responsible for myocarditis by secreting proinflammatory cytokines. The picture is further muddled when considering the polarization of macrophages into a pro-inflammatory M1 phase versus an anti-inflammatory M2 phase. M1 macrophages are associated with worse myocarditis, whereas M2 macrophages reduce myocarditis (70). Similarly, NK cells inhibit viral replication (71) and ameliorate myocarditis (72), primarily via secretion of IFN γ (73). Together, macrophages and NK cells quickly clear infected cells and poise the environment for adaptive responses.

The cytokines and chemokines secreted by the innate immune cells attract and drive differentiation of adaptive immune cells. CVB3 infection induces CD4⁺ helper T (Th) and CD8⁺ cytotoxic T (Tc) responses in the heart. Mice deficient for either have increased viral titers (74). Th cells differentiate into four main subtypes: proinflammatory Th1 and Th17 cells and anti-inflammatory Th2 and regulatory T (Treg) cells. Th1 cells arise in the presence of type I/II IFNs and activate Tc cells (75). Th1 cells also produce large amounts of IFN γ , which promotes antibody class-switching in B cells to promote humoral immunity. Th17

cells arise in the presence of IL-6 and TGF β (75). Upon infection with CVB3, Th17 cells increase during the first 10 days of infection, while Th1 cells peak on days 5-7 in the heart (76). Both exacerbate myocarditis and promote Tc activation. Tc cells contain a unique T cell receptor (TCR) that can recognize foreign material displayed on the surface of tissue cells by the major histocompatibility complex-I, which is the signal to kill the cell (77). Upon CVB3 infection, CD8⁺ T cells peak at days 5-7 to clear infected cells (76). To limit an overactive immune response, Treg cells provide negative feedback that reduces inflammation (78), closing the loop on T cell responses.

The role of B cells is determined by its contribution to the cytokine milieu. IL-10 producing B cells limit Th1 and Th17 cells and reduce inflammation (79). Alternatively, B cells producing TNF- α , IL-6, and IL-17 promote Th1 and Th17 differentiation while preventing Th2 differentiation, which serves to promote inflammation (80). B cells also promote formation of Treg cells in the heart following CVB3 infection (81). Thus, it is difficult to define a singular contribution of B cells to myocarditis. Altogether, the literature suggests that the cytokine/chemokine-coordinated recruitment of innate- and adaptive-immune infiltrates modules myocarditis.

Viral myocarditis has been long known to depend on both CVB3 cardiotoxicity and a proinflammatory immune response. Other active areas of research involve the role of microRNA (82), long noncoding RNA (83), and inflammasomes (84). Further, sex and host genetics play a role in the severity of myocarditis (19, 20). Frankly, we are at the point where a mechanistic understanding of how signaling networks, immunity, and viral cytotoxicity produce myocarditis—all conditional on host genetics—will need computational approaches.

1.4 Systems virology

Systems biology is an approach that—to put it colloquially—seeks to understand how the whole is greater than the sum of its parts, with a goal to generate testable hypotheses or discover emergent properties. To do so, the individual pieces (e.g., proteins in a signaling network) are organized through mathematical encoding of their relationships (model). The mathematical solution provides some prediction about the system (mine), which is used to design an experiment (manipulate). The results (measure) may shed light on the pieces of the model that need further refinement, such as an updated parameter or insertion of a new species into the model (model again). This iterative process of model → mine → manipulate → measure is the mantra of systems biology. The point at which the model is considered “validated” and stops being updated is difficult to define and specific to the field of research or biological question. Regardless, once “validated,” the model is

useful for examining the emergent properties of the system—or, the behavior of the system that is observed by the interactions of the individual pieces. This process has been widely used in studying cardiovascular signaling networks (85), metabolism (86), and cancer (87).

Systems biology has had less traction in the virology community despite these approaches beginning decades ago with mathematical models of bacteriophage T4 (88–90). Systems virology gained traction in the 1990s with models of HIV-1 infection, such as a classic study by Perelson et al. that simulated the dynamics of infection and viral clearance (91). Susceptible target cells were “infected” at a constant rate and could recover or produce new virus, which was added to the pool of free virus to infect target cells. Similar models were made for hepatitis B and C viruses around the same time (92, 93). Since then, model elaborations have included immune responses and have expanded to viruses such as influenza, dengue virus, and ebolavirus (32). These models were useful for determining the efficacy of drugs and at which phases of infection the dose is most appropriate.

Systems-level measurements of host responses to infection result in large datasets, such as from high-throughput assays (28, 94) or transcriptomics (95). To draw conclusions from these data, statistical models are used to organize the samples by similarity (clustering), reduce the dimensionality into a more-interpretable set of variables (e.g., principal component analysis), or establish input–output relationships among variables (e.g., partial least squares regression) (96). Once simplified, mechanisms can be unpackaged experimentally (97). Transcriptomics is commoditized now, and well-documented computational packages exist that make clustering analysis accessible for non-computationalists. Other statistical models are less common in virology but have proved useful for determining signaling networks hijacked by CVB3 to promote cell death (28, 94). Now in the era of big data, statistical models will become increasingly useful for virologists seeking to understand virus–host interactions.

Mechanistic models relate species through chemical rate equations that encode enzyme activity, compartment-switching (e.g., nuclear vs. cytoplasmic), and/or synthesis–degradation (97). Rate parameters are measured by biochemical studies. In cases where rate constants are unknown, parameters are estimated heuristically to fit model predictions to validation data. Mechanistic models are strongly suited for identifying feedbacks and crosstalk important for capturing the dynamics of a system (97). Such models have emerged to study facets of the viral life cycle, but many lack a holistic integration of the entire life cycle (7). Mechanistic modeling of viruses is an active area of research, with models recently emerging for hepatitis C virus (98) and dengue virus (99). In virology, mechanistic models are useful because one can tune parameters to reflect different host-cell conditions and make predictions about how the trajectory of infection will be altered.

1.5 Overview of dissertation

In this dissertation, I seek to elucidate how population-level differences in host-cell species alter enteroviral infection and resulting disease. A study of how changes in host-cell proteins will affect the dynamics of CVB3 infection requires the use of systems-biology approaches. To achieve this goal, mechanistic and statistical modeling will be used. Mechanistic modeling is desirable for tracking dynamics of viral and host species and can be used to generate hypotheses. To identify population-wide differences in key host proteins, statistical modeling will be used to obtain parameters for the mechanistic model. With a robust computational toolset, the daunting task of identifying critical virus–host interactions becomes achievable.

In chapter 2, I describe the development of a mechanistic model for the complete kinetics of CVB3. We constructed the model with a modular design, with each module encoding a stage of the viral life cycle and experimentally validated. We included a generalized host-cell innate immune response as a series of negative feedbacks overlaid on the viral life cycle. Importantly, the model is constructed with little parameter fitting. We generate a testable hypothesis surrounding the emergent properties of the system, leading us to a bioinformatic search into the population genetics of the protein MAVS. Ultimately, we identify a crucial MAVS polymorphism that is able to alter susceptibility to CVB3 infection.

In chapter 3, I describe the conversion of the command-line CVB3 model into a graphical user interface. This activity was spurred by an invited submission to *STAR Protocols*. The hope is that an accessible model of viral infection will encourage non-computationalists to employ models in their research or classrooms.

In chapter 4, I personalize the CVB3 model to 1489 individuals by using publicly available data to alter parameters that are variable within the human population. However, to use this data, I had to first develop a tool to estimate protein abundances from transcriptomic data. I combined publicly available transcriptomic and proteomic data from the Cancer Cell Line Encyclopedia and generated statistical models for 4000+ genes that related mRNA abundance to absolute protein copy numbers. We examined characteristics of these relationships and validated protein copy-number estimations against experimental proteomic data. Finally, we apply the tool to estimation of receptor abundances for CVB3 to generate the personalized models. Since the tool was more generalizable, we also used it to estimate the proteome of 796 breast cancers. We examined how the molecular classification differed between clustering by transcriptomics and proteomics. Then, we honed in on the genes that drive these reclassifications. Interestingly, we find many genes of clinical and therapeutic relevance to breast cancer.

In chapter 5, I discuss the broader impacts, limitations, and future directions for the work presented in chapters 2 and 4. Further, I will present data that did not fit nicely into a chapter of this dissertation but that were major activities during my PhD. I package those data and conclusions that can be drawn from them, as well as indicate where future work is needed.

Ultimately, my dissertation identifies two proteins—MAVS and CAR—that are important for defining susceptibility to CVB3 disease.

1.6 References

1. K. J. Olival, *et al.*, Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
2. S. T. Jacob, *et al.*, Ebola virus disease. *Nat Rev Dis Primers* **6**, 13 (2020).
3. F. Krammer, *et al.*, Influenza. *Nat Rev Dis Primers* **4**, 3 (2018).
4. J. Puenpa, N. Wanlapakorn, S. Vongpunsawad, Y. Poovorawan, The History of Enterovirus A71 Outbreaks and Molecular Epidemiology in the Asia-Pacific Region. *J Biomed Sci* **26**, 75 (2019).
5. S. E. Crawford, *et al.*, Rotavirus infection. *Nat Rev Dis Primers* **3**, 17083 (2017).
6. S. E. Jacobs, D. M. Lamson, K. St George, T. J. Walsh, Human rhinoviruses. *Clin Microbiol Rev* **26**, 135–162 (2013).
7. J. Yin, J. Redovich, Kinetic Modeling of Virus Growth in Cells. *Microbiol Mol Biol Rev* **82** (2018).
8. Connecting the dots from viral infection to disease. *Nat Microbiol* **8**, 1363–1364 (2023).
9. P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155–170 (2021).
10. P. V. Markov, *et al.*, The evolution of SARS-CoV-2. *Nat Rev Microbiol* **21**, 361–379 (2023).
11. COVID-19 Host Genetics Initiative, Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
12. A. Onoja, *et al.*, An explainable model of host genetic interactions linked to COVID-19 severity. *Commun Biol* **5**, 1133 (2022).
13. I.-E. Galani, *et al.*, Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat Immunol* **22**, 32–40 (2021).

14. S. Mohandas, *et al.*, Immune mechanisms underlying COVID-19 pathology and post-acute sequelae of SARS-CoV-2 infection (PASC). *eLife* **12**, e86014.
15. M. Yin-Murphy, J. W. Almond, "Picornaviruses" in *Medical Microbiology*, 4th Ed., S. Baron, Ed. (University of Texas Medical Branch at Galveston, 1996) (July 23, 2020).
16. V. Racaniello, "Picornaviridae: the viruses and their replication" in *Fields Virology*, 6th ed, D. M. Knipe, P. M. Howley, Eds. (Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013).
17. J. F. Enders, T. H. Weller, F. C. Robbins, Cultivation of the Lansing Strain of Poliomyelitis Virus in Cultures of Various Human Embryonic Tissues. *Science* **109**, 85–87 (1949).
18. M. Science, U. Allen, Enterovirus D68 and disease severity: more questions than answers. *CMAJ* **187**, 1269–1270 (2015).
19. N. Lasrado, J. Reddy, An overview of the immune mechanisms of viral myocarditis. *Rev Med Virol* **30**, 1–14 (2020).
20. L. J. Wolfgram, K. W. Beisel, A. Herskowitz, N. R. Rose, Variations in the susceptibility to Coxsackievirus B3-induced myocarditis among different strains of mice. *The Journal of Immunology* **136**, 1846–1852 (1986).
21. K. Adachi, A. Muraishi, Y. Seki, K. Yamaki, M. Yoshizuka, Coxsackievirus B3 genomes detected by polymerase chain reaction: evidence of latent persistency in the myocardium in experimental murine myocarditis. *Histol Histopathol* **11**, 587–596 (1996).
22. P. K.-M. Cheung, *et al.*, Specific interactions of mouse organ proteins with the 5'untranslated region of coxsackievirus B3: potential determinants of viral tissue tropism. *J Med Virol* **77**, 414–424 (2005).
23. F. S. Garmaroudi, *et al.*, Coxsackievirus B3 replication and pathogenesis. *Future Microbiol* **10**, 629–653 (2015).
24. Y. Shi, *et al.*, Cardiac deletion of the Coxsackievirus-adenovirus receptor abolishes Coxsackievirus B3 infection and prevents myocarditis in vivo. *J Am Coll Cardiol* **53**, 1219–1226 (2009).
25. Y. Seko, N. Takahashi, H. Yagita, K. Okumura, Y. Yazaki, Expression of cytokine mRNAs in murine hearts with acute myocarditis caused by coxsackievirus b3. *J Pathol* **183**, 105–108 (1997).
26. G. Fung, H. Luo, Y. Qiu, D. Yang, B. McManus, Myocarditis. *Circ Res* **118**, 496–514 (2016).
27. T. Yajima, K. U. Knowlton, Viral myocarditis: from the perspective of the virus. *Circulation* **119**, 2615–2624 (2009).

28. F. S. Garmaroudi, *et al.*, Pairwise network mechanisms in the host signaling response to coxsackievirus B3 infection. *Proc Natl Acad Sci U S A* **107**, 17053–17058 (2010).
29. A. J. Sadler, B. R. G. Williams, Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559–568 (2008).
30. J. W. Schoggins, *et al.*, A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
31. J. Zhu, *et al.*, Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**, 936–948 (2014).
32. C. Zitzmann, L. Kaderali, Mathematical Analysis of Viral Replication Dynamics and Antiviral Treatment Strategies: From Basic Models to Age-Based Multi-Scale Modeling. *Front Microbiol* **9**, 1546 (2018).
33. D. R. Shafren, D. T. Williams, R. D. Barry, A decay-accelerating factor-binding strain of coxsackievirus B3 requires the coxsackievirus-adenovirus receptor protein to mediate lytic infection of rhabdomyosarcoma cells. *J Virol* **71**, 9844–9848 (1997).
34. C. B. Coyne, J. M. Bergelson, Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. *Cell* **124**, 119–131 (2006).
35. J. M. Bergelson, *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320–1323 (1997).
36. C. J. Cohen, *et al.*, The coxsackievirus and adenovirus receptor is a transmembrane component of the tight junction. *Proc Natl Acad Sci U S A* **98**, 15191–15196 (2001).
37. D. Yang, *et al.*, In vitro mutational and inhibitory analysis of the cis-acting translational elements within the 5' untranslated region of coxsackievirus B3: potential targets for antiviral action of antisense oligomers. *Virology* **228**, 63–73 (1997).
38. N.-Y. Hsu, *et al.*, Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* **141**, 799–811 (2010).
39. R. W. A. L. Limpens, *et al.*, The transformation of enterovirus replication structures: a three-dimensional study of single- and double-membrane compartments. *mBio* **2**, e00166-11 (2011).
40. C. Netherton, K. Moffat, E. Brooks, T. Wileman, A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. *Adv Virus Res* **70**, 101–182 (2007).
41. S. Miller, J. Krijnse-Locker, Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**, 363–374 (2008).
42. A. V. Paul, J. H. van Boom, D. Filippov, E. Wimmer, Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* **393**, 280–284 (1998).

43. M. J. M. van Ooij, *et al.*, Structural and functional characterization of the coxsackievirus B3 CRE(2C): role of CRE(2C) in negative- and positive-strand RNA synthesis. *J Gen Virol* **87**, 103–113 (2006).
44. J. E. Novak, K. Kirkegaard, Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. *J Virol* **65**, 3384–3387 (1991).
45. P. Jiang, Y. Liu, H.-C. Ma, A. V. Paul, E. Wimmer, Picornavirus morphogenesis. *Microbiol Mol Biol Rev* **78**, 418–437 (2014).
46. Y. Liu, *et al.*, Direct Interaction between Two Viral Proteins, the Nonstructural Protein 2CATPase and the Capsid Protein VP3, Is Required for Enterovirus Morphogenesis. *PLoS Pathog* **6**, e1001066 (2010).
47. D. Etchison, S. C. Milburn, I. Edery, N. Sonenberg, J. W. Hershey, Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 220,000-dalton polypeptide associated with eucaryotic initiation factor 3 and a cap binding protein complex. *J Biol Chem* **257**, 14806–14810 (1982).
48. S. Wu, *et al.*, Protease 2A induces stress granule formation during coxsackievirus B3 and enterovirus 71 infections. *Virol J* **11**, 192 (2014).
49. N. M. Kuyumcu-Martinez, M. E. Van Eden, P. Younan, R. E. Lloyd, Cleavage of poly(A)-binding protein by poliovirus 3C protease inhibits host cell translation: a novel mechanism for host translation shutoff. *Mol Cell Biol* **24**, 1779–1790 (2004).
50. Q. Feng, *et al.*, Enterovirus 2Apro targets MDA5 and MAVS in infected cells. *J Virol* **88**, 3369–3378 (2014).
51. A. Mukherjee, *et al.*, The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog* **7**, e1001311 (2011).
52. B. Wang, *et al.*, Enterovirus 71 protease 2Apro targets MAVS to inhibit anti-viral type I interferon responses. *PLoS Pathog* **9**, e1003231 (2013).
53. C. Badorff, *et al.*, Enteroviral protease 2A directly cleaves dystrophin and is inhibited by a dystrophin-based substrate analogue. *J Biol Chem* **275**, 11191–11197 (2000).
54. G. A. Belov, *et al.*, Early alteration of nucleocytoplasmic traffic induced by some RNA viruses. *Virology* **275**, 244–248 (2000).
55. C. T. Cornell, W. B. Kiosses, S. Harkins, J. L. Whitton, Inhibition of protein trafficking by coxsackievirus b3: multiple viral proteins target a single organelle. *J Virol* **80**, 6637–6647 (2006).

56. G. A. Belov, *et al.*, Bidirectional increase in permeability of nuclear envelope upon poliovirus infection and accompanying alterations of nuclear pores. *J Virol* **78**, 10166–10177 (2004).
57. T. Kawai, *et al.*, IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol* **6**, 981–988 (2005).
58. E. Meylan, *et al.*, Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
59. R. B. Seth, L. Sun, C.-K. Ea, Z. J. Chen, Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).
60. L.-G. Xu, *et al.*, VISA is an adapter protein required for virus-triggered IFN-beta signaling. *Mol Cell* **19**, 727–740 (2005).
61. H. Kato, K. Takahashi, T. Fujita, RIG-I-like receptors: cytoplasmic sensors for non-self RNA. *Immunol Rev* **243**, 91–98 (2011).
62. F. Hou, *et al.*, MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146**, 448–461 (2011).
63. Z. Cao, J. Xiong, M. Takeuchi, T. Kurama, D. V. Goeddel, TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
64. V. Baud, *et al.*, Signaling by proinflammatory cytokines: oligomerization of TRAF2 and TRAF6 is sufficient for JNK and IKK activation and target gene induction via an amino-terminal effector domain. *Genes Dev* **13**, 1297–1308 (1999).
65. R. Fang, *et al.*, MAVS activates TBK1 and IKK ϵ through TRAFs in NEMO dependent and independent manner. *PLoS Pathog* **13**, e1006720 (2017).
66. S. Liu, *et al.*, Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, aaa2630 (2015).
67. D. Kang, *et al.*, mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A* **99**, 637–642 (2002).
68. C. D. Gregory, A. Devitt, The macrophage and the apoptotic cell: an innate immune interaction viewed simplistically? *Immunology* **113**, 1–14 (2004).
69. C. Bogdan, Nitric oxide and the immune response. *Nat Immunol* **2**, 907–916 (2001).
70. K. Li, *et al.*, Differential macrophage polarization in male and female BALB/c mice infected with coxsackievirus B3 defines susceptibility to viral myocarditis. *Circ Res* **105**, 353–364 (2009).
71. J. Yuan, *et al.*, CXCL10 inhibits viral replication through recruitment of natural killer cells in coxsackievirus B3-induced myocarditis. *Circ Res* **104**, 628–638 (2009).

72. K. Klingel, *et al.*, The activating receptor NKG2D of natural killer cells promotes resistance against enterovirus-mediated inflammatory cardiomyopathy. *J Pathol* **234**, 164–177 (2014).
73. M. H. Hühn, *et al.*, IFN-gamma production dominates the early human natural killer cell response to Coxsackievirus infection. *Cell Microbiol* **10**, 426–436 (2008).
74. A. Henke, S. Huber, A. Stelzner, J. L. Whitton, The role of CD8⁺ T lymphocytes in coxsackievirus B3-induced myocarditis. *J Virol* **69**, 6720–6728 (1995).
75. S. L. Swain, K. K. McKinstry, T. M. Strutt, Expanding roles for CD4⁺ T cells in immunity to viruses. *Nat Rev Immunol* **12**, 136–148 (2012).
76. J. Yuan, *et al.*, Th17 cells contribute to viral replication in coxsackievirus B3-induced acute viral myocarditis. *J Immunol* **185**, 4004–4010 (2010).
77. B. Alberts, *et al.*, “T Cells and MHC Proteins” in *Molecular Biology of the Cell. 4th Edition*, (Garland Science, 2002) (August 22, 2023).
78. Y. Shi, *et al.*, Regulatory T cells protect mice against coxsackievirus-induced myocarditis through the transforming growth factor beta-coxsackie-adenovirus receptor pathway. *Circulation* **121**, 2624–2634 (2010).
79. B. Wei, Y. Deng, Y. Huang, X. Gao, W. Wu, IL-10-producing B cells attenuate cardiac inflammation by regulating Th1 and Th17 cells in acute viral myocarditis induced by coxsackie virus B3. *Life Sci* **235**, 116838 (2019).
80. Z. Cen, *et al.*, The Role of B Cells in Regulation of Th Cell Differentiation in Coxsackievirus B3-Induced Acute Myocarditis. *Inflammation* **44**, 1949–1960 (2021).
81. J. Lu, *et al.*, The absence of B cells disrupts splenic and myocardial Treg homeostasis in coxsackievirus B3-induced myocarditis. *Clin Exp Immunol* **208**, 1–11 (2022).
82. M. F. Corsten, *et al.*, MicroRNA profiling identifies microRNA-155 as an adverse mediator of cardiac injury and dysfunction during acute viral myocarditis. *Circ Res* **111**, 415–425 (2012).
83. X. Nie, *et al.*, Expression Profiles and Potential Functions of Long Non-Coding RNAs in the Heart of Mice With Coxsackie B3 Virus-Induced Myocarditis. *Front Cell Infect Microbiol* **11**, 704919 (2021).
84. K. Miteva, *et al.*, Mesenchymal stromal cells inhibit NLRP3 inflammasome activation in a model of Coxsackievirus B3-induced inflammatory cardiomyopathy. *Sci Rep* **8**, 2820 (2018).
85. K. A, *et al.*, Signaling network model of cardiomyocyte morphological changes in familial cardiomyopathy. *Journal of molecular and cellular cardiology* **174** (2023).

86. E. Brunk, *et al.*, Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272–281 (2018).
87. E. J. Pereira, *et al.*, Sporadic activation of an oxidative stress-dependent NRF2-p53 signaling network in breast epithelial spheroids and premalignancies. *Sci Signal* **13**, eaba4200 (2020).
88. N. A. Barricelli, L. Nelson, R. Toombs, Virus-genetic theory testing by data processing machines. I. Basic procedures and models tested. *J Theor Biol* **32**, 559–580 (1971).
89. N. A. Barricelli, R. Toombs, Virus-genetic theory testing by data processing machines. II. Fit of classical genetic T4 data. *J Theor Biol* **32**, 581–601 (1971).
90. N. A. Barricelli, R. Toombs, Virus-genetic theory testing by data processing machines. 3. The interpretation of radiation genetic and partial phage experiments in T4. General remarks on the theory-testing results. *J Theor Biol* **32**, 603–623 (1971).
91. A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, D. D. Ho, HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996).
92. A. U. Neumann, *et al.*, Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* **282**, 103–107 (1998).
93. M. A. Nowak, *et al.*, Viral dynamics in hepatitis B virus infection. *Proc Natl Acad Sci U S A* **93**, 4398–4402 (1996).
94. K. J. Jensen, *et al.*, An ERK-p38 subnetwork coordinates host cell apoptosis and necrosis during coxsackievirus B3 infection. *Cell Host Microbe* **13**, 67–76 (2013).
95. M. J. Korth, G. L. Law, Systems Virology. *Viral Pathogenesis*, 141–156 (2016).
96. K. A. Janes, M. B. Yaffe, Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* **7**, 820–828 (2006).
97. K. A. Janes, D. A. Lauffenburger, Models of signalling networks - what cell biologists can gain from them and give to them. *J Cell Sci* **126**, 1913–1921 (2013).
98. M. Binder, *et al.*, Replication vesicles are load- and choke-points in the hepatitis C virus lifecycle. *PLoS Pathog* **9**, e1003561 (2013).
99. C. Zitzmann, *et al.*, A Coupled Mathematical Model of the Intracellular Replication of Dengue Virus and the Host Cell Immune Response to Infection. *Front Microbiol* **11**, 725 (2020).

Chapter 2:

Modeling the complete kinetics of coxsackievirus B3 reveals human determinants of host-cell feedback[#]

Aaron B. Lopacinski,^{1,4} Andrew J. Sweatt,^{1,4} Christian M. Smolko,¹
Elise Gray-Gaillard,^{1,2} Cheryl A. Borgman,¹ Millie Shah,^{1,*} and
Kevin A. Janes,^{1,3,5,*}

¹Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

²Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

³Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, USA

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: ms2kf@virginia.edu (M.S.), kjanes@virginia.edu (K.A.J.)

[#]This work has been published in *Cell Systems* and is available on the ScienceDirect web site: <https://doi.org/10.1016/j.cels.2021.02.004>.

Supplementary tables are too large to be included in this dissertation. Please refer to the source publication to obtain them.

2.1 Abstract

Complete kinetic models are pervasive in chemistry but lacking in biological systems. We encoded the complete kinetics of infection for coxsackievirus B3 (CVB3), a compact and fast-acting RNA virus. The kinetics are built from detailed modules for viral binding–delivery, translation–replication, and encapsidation. Specific module activities are dampened by the type I interferon response to viral double-stranded RNAs (dsRNAs), which is itself disrupted by viral proteinases. The validated kinetics uncovered that cleavability of the dsRNA transducer mitochondrial antiviral signaling protein (MAVS) becomes a stronger determinant of viral outcomes when cells receive supplemental interferon after infection. Cleavability is naturally altered in humans by a common MAVS polymorphism, which removes a proteinase-targeted site but paradoxically elevates CVB3 infectivity. These observations are reconciled with a simple nonlinear model of MAVS regulation. Modeling complete kinetics is an attainable goal for small, rapidly infecting viruses and perhaps viral pathogens more broadly.

2.2 Introduction

In chemistry, a complete kinetic model specifies products and reactants, their stoichiometry, any catalysts, and the step-by-step sequence of key elementary reactions from start to finish (1). Direct analogies in biological systems are hard to identify because the steps are uncertain and the definition of “start” or “finish” is often arbitrary. Not so for picornaviruses, a family of non-enveloped, positive-strand RNA viruses that infect, rapidly amplify their genetic material and capsid proteins, and lyse the host cell marking termination. Steps in the picornaviral life cycle are fully delineated (2). The translation of the coding RNA genome as a single polypeptide ensures an equal proportion of each viral protein during synthesis. From the standpoint of products and reactants, picornaviruses present an opportunity to construct complete kinetic models.

The power of kinetic mechanisms lies in the explanatory and predictive models that they generate (1). Mathematical models of viral translation–replication have a rich history, but they are formally incomplete in omitting the details of early binding–entry, intermediate antiviral pathways, and late self-assembly of viral particles (3). Some omissions are for simplicity and scope, but others are for lack of biological parameters or quantities. As human disease agents, picornaviruses within the *Enterovirus* genus—which includes rhinovirus, poliovirus, and coxsackievirus strains—have been parameterized extensively (2). Enteroviruses thus represent the best-case testbed for proposing a complete kinetic model of a biological process if one can be defined.

Here, we drafted the complete kinetics for coxsackievirus B3 (CVB3), an enterovirus

that infects the heart and drives viral myocarditis (4). With attention to molecular detail, we encoded interacting modules for viral delivery, translation–replication, and encapsidation, overlaying a prototypical antiviral response along with viral antagonism of that response (Figure 2.1A). Each component presented its own challenges, which we tackled module-by-module before simulating the complete kinetics of CVB3 infection. In a cardiomyocyte-derived cell line, the integrated modules captured host-cell susceptibility and viral RNA–protein dynamics with little-to-no parameter fitting of specific mechanisms. The generalized antiviral and antagonistic feedbacks were screened combinatorially to identify time-dependent interactions later verified experimentally. Host and viral feedbacks converged on an important role for MAVS, a polymorphic host-cell transducer of the antiviral response to dsRNA that is cleaved by an enterovirus-encoded proteinase (5). The centrality of MAVS led to the discovery of a common variant in humans (MAVS Q93E), which reroutes cleavage during CVB3 infection and favors viral propagation. The net effect of this polymorphism on feedback arises from nonlinearities in MAVS regulation, as illustrated by a standalone model of its provisional activation mechanism. For virology, complete kinetics provide a systems-level end goal for decades of research and possibly an aid to prioritizing scientific activity during viral outbreaks.

2.3 Results

2.3.1 Modular draft of a complete kinetic model for the CVB3 life cycle

We pursued a complete kinetic model for acute CVB3 infection by deconstructing its life cycle into separable modules that could be developed independently (Figure 2.1A). The cellular tropism of CVB3 is determined by specific cell-surface receptors (6, 7), which internalize the virus before endosomal escape of its positive-strand RNA genome (R_p) into the cytoplasm. From here, the intracellular steps generalize to all enteroviruses: i) translation of R_p into polyprotein ($CVB3_{pp}$), ii) maturation of $CVB3_{pp}$ into capsid subunits and nonstructural proteins for the virus, iii) cis replication of R_p through a negative-strand intermediate by RNA-dependent RNA polymerase ($3D^{pol}$) localized to host-cell membranes hijacked by the virus, and iv) encapsidation of R_p around 12 capsid pentamers aided by hydrophobic proteins ($2C^{ATPase}$) and concluding with lytic release from the host (8). In contrast to cell-signaling pathways (9), the viral conduits to and from these modules are clearly defined, and an acute infection usually completes within hours if unimpeded.

At the nexus of the enteroviral life cycle is the type I interferon response. Viral replication is associated with long, double-stranded replicative intermediates of CVB3 (dsCVB3) that are sensed by innate antiviral pathways like other dsRNAs [e.g., poly(I:C)] (10). Cytosolic dsCVB3 gives rise to induction of *IFNB*, autocrine–paracrine signaling, and

the induction of interferon-stimulated genes (ISGs) (11), such as oligoadenylate synthetases and Mx-family GTPases (Figure 2.1B). There are hundreds of different ISGs, several of which inhibit core enteroviral processes at multiple points (Figure 2.1A) (12). ISGs are predicted to persist for the duration of an infection (Figure 2.1C), and active type I interferon signaling is a potent inhibitor of enteroviral replication in vivo (13, 14). The enteroviral proteinases ($2A^{pro}$ – $3C^{pro}$)—which mature the $CVB3_{pp}$ polyprotein and steal host-cell ribosomes for cap-independent viral translation—interfere with activation of the interferon response by cleavage of dsRNA sensors and transducers (Figure 2.1A) (5, 15, 16). Cleavage of host-cell proteins occurs on the same time scale as polyprotein maturation (Figure 2.1D) (5, 17), requiring us to overlay a network of ISG-related negative feedbacks on the CVB3 life cycle (Figure 2.1A, purple).

The viral life-cycle modules and antiviral–antagonistic feedbacks were encoded as a system of 54 differential equations that are derived in Section 2.5, STAR Methods. Experimental evidence for 91% of the 92 parameters is provided in Table S1. In general engineering design, modules are appealing because they enable individual components to be developed and characterized before integration (18). Accordingly, our results are communicated to retain the modular organization of the CVB3 model. For individual modules, we expand biochemical mechanisms, describe critical assumptions or considerations, compare with experiments, and conclude with non-obvious computational predictions.

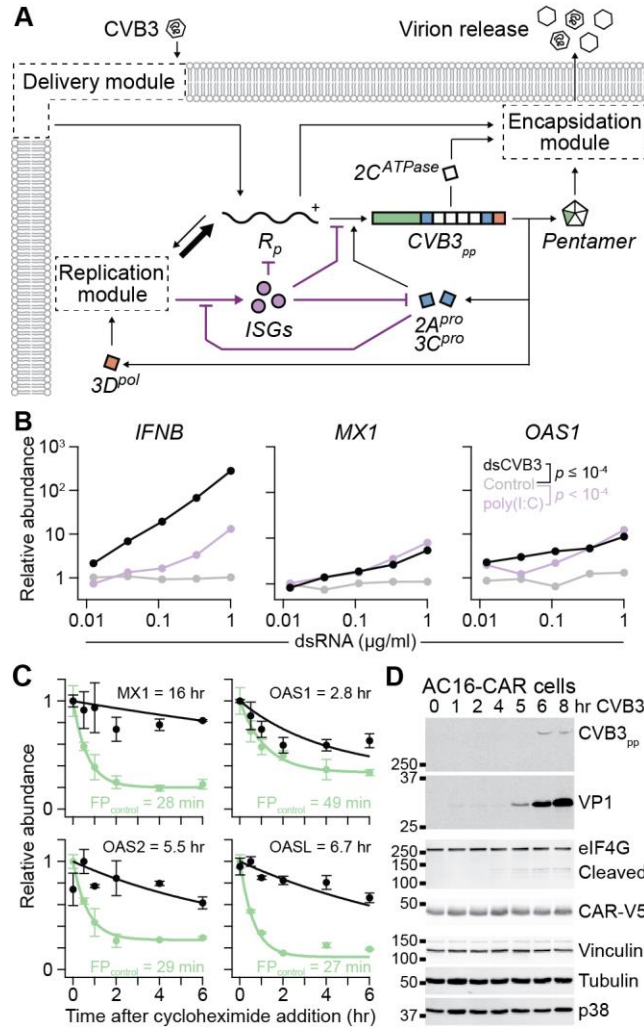


Figure 2.1. A modular encoding of the CVB3 life cycle elaborated with antiviral responses and viral antagonism of host-cell processes.

(A) Overview of the CVB3 model architecture. CVB3 enters through a module of receptors and trafficking states that delivers its positive-strand RNA genome (R_p) to the cytoplasm. R_p is translated into a polyprotein ($CVB3_{pp}$) that matures into capsid subunits contributing one fifth of a subunit (Pentamer) in a 12-subunit assembly, enteroviral proteinases ($2A^{pro}$ – $3C^{pro}$), hydrophobic proteins ($2C^{ATPase}$), and RNA-dependent RNA polymerase ($3D^{pol}$). $3D^{pol}$ replicates R_p through a negative-strand intermediate in a module that gives rise to excess R_p , which joins with Pentamer in an encapsidation module that self-assembles Pentamer around R_p and leads to virion release. Interferon stimulated genes (ISGs) are induced as a consequence of the dsRNA associated with CVB3 replication and impede viral progression where indicated; viral sensing is also antagonized by $3C^{pro}$ (5). Expanded descriptions of the delivery, replication, and encapsidation modules are shown in Figures 2.3A, 2.5A, 2.7A, and 2.7D, and model parameters are available in Table S1.

(B) Transfected dsCVB3 elicits a robust type I interferon response. AC16-CAR cells were lipofected with $n = 5$ doses of the indicated dsRNAs for four hours and analyzed for the indicated ISGs with *PRDX6*, *HINT1*, and *GUSB* used for loading normalization. Differences between conditions were assessed by Šidák-corrected, log-transformed three-way ANOVA. Uninduced target genes (*EIF2AK2* and *GAPDH*) are analyzed in Figure 2.2A.

(C) ISGs are long lived when translated. 293T/17 cells were lipofected with V5 epitope-tagged plasmids, treated with 50 μ M cycloheximide for the indicated times, and lysed for quantitative immunoblotting. Half-lives for the ISGs MX1, OAS1, OAS2, OASL, and FP_{control} (a fast-degrading protein fragment used as a positive

control) were estimated by nonlinear least-squares curve fitting. Data are shown as the mean \pm range of biological duplicates at $n = 6$ different time points. Representative immunoblots are shown in Figure 2.2B.

(D) CVB3 polyprotein maturation coincides with cleavage of host-cell targets. AC16-CAR cells were infected with CVB3 at MOI = 10 for the indicated times and immunoblotted for VP1 capsid protein (also present in the full-length polyprotein [CVB3_{pp}]) and eIF4G [a host protein cleaved by 2A^{pro} (16)] with ectopic CAR-V5, vinculin, tubulin, and p38 used as loading controls. The image gamma was adjusted for polyprotein–VP1 (gamma = 20) and eIF4G (gamma = 2) to show band pairs at the same exposure.

See also Figure 2.2 and Table S1.

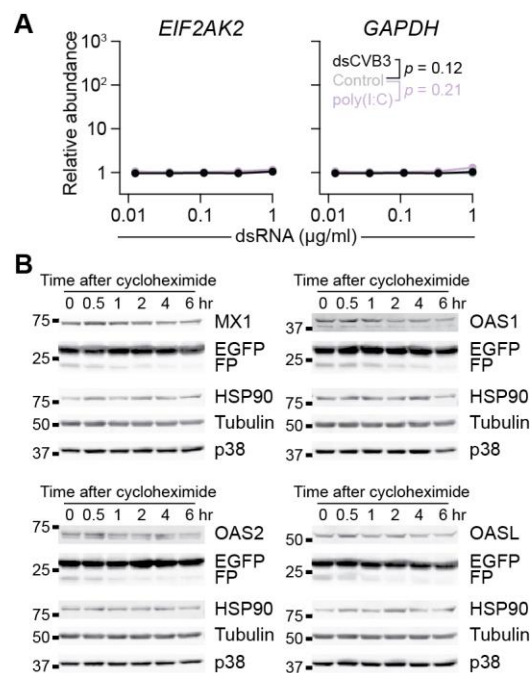


Figure 2.2. Control genes for dsRNA induction and representative immunoblots of the cycloheximide experiment quantified in Figure 2.1C.

(A) dsRNA lipofection does not give rise to nonspecific gene induction. AC16-CAR cells were lipofected with $n = 5$ doses of the indicated dsRNAs for four hours and analyzed for the indicated genes with *PRDX6*, *HINT1*, and *GUSB* used for loading normalization. Differences between conditions were assessed by Šidák-corrected, log-transformed three-way ANOVA.

(B) Representative immunoblots used for half-life quantification. 293T/17 cells were lipofected with V5 epitope-tagged plasmids (equal mixtures of ISG and EGFP), treated with 50 µM cycloheximide for the indicated times, and lysed for quantitative immunoblotting. A C-terminal EGFP fragment (FP) was used as a control for a fast-degrading protein.

2.3.2 Explanatory modeling of CVB3 tropism requires careful extracellular bookkeeping

A hurdle to modeling the complete kinetics of CVB3 arises even before the virus has encountered a host cell. Stock preparations of enterovirus contain many more RNA-filled viral particles than infectious plaque-forming-units (PFUs). Encapsidated RNA genomes often do not reach a host-cell ribosome for translation, and some of those that do contain deleterious mutations from the prior replication. These particles are “defective” from the standpoint of infection, yet they are generally able to bind cell-surface receptors and internalize (19). Defective particles thus contribute viral protein and RNA to the system, which must be accounted for at the start of the life cycle.

Particle-to-PFU ratios vary widely within an enterovirus species—the ratio for poliovirus, for example, is reported to be 30:1 to 1000:1 (2). We quantified sedimentable RNA content from separately purified CVB3 preparations of known PFU titer and estimated a particle-to-PFU ratio of 800 ± 200 ($n = 4$ preparations; Section 2.5, STAR Methods). In the model, we assumed an 800-fold excess of defective particles, which traffic identically to PFUs but do not translate upon entering the host-cell cytoplasm (Section 2.5, STAR Methods). The assumption creates a kinetic dead-end for defective particles without any additional kinetic parameters (Table S1). More complicated defective-particle fates yielded similar infection outcomes, as shown after the model was fully developed (see below).

A second challenge relates to the front-end model implementation of viral delivery. The early steps of virus binding and internalization are discrete and thus intrinsically stochastic (20). However, one negative-strand template can yield hundreds of protein-coding positive strands within an hour (21), indicating rapid transition to a regime where deterministic modeling is valid. We balanced these tradeoffs by postulating that the dominant stochasticity was the Poisson noise from the number of PFUs encountered at a given multiplicity of infection (MOI) describing the average PFU per cell within a population (20). Cell-to-cell variation in all downstream processes was modeled by lognormal random variables centered around the best estimates from the literature (Table S1) and sampled with a user-defined coefficient of variation. We configured the simulations to run in one of two modes: 1) a “single-cell” mode, in which a discrete number of PFUs is simulated with lognormally distributed downstream parameters, and 2) a “cell-population” mode, in which a lognormal instance of the life cycle is paired with a PFU integer drawn from a Poisson distribution about a continuous MOI representing the average. In either mode, dynamic trajectories were well summarized by ~100 separate iterations. The dual implementation thereby models the average-cell response to an initial condition that is highly stochastic.

Binding and delivery of CVB3 requires decay accelerating factor (DAF, officially named CD55) and coxsackievirus and adenovirus receptor (CAR, officially named CXADR) (6, 7). DAF is a low-affinity receptor ($K_D \sim 3 \mu\text{M}$) whose transcript is widely expressed at moderate abundance in human tissues (median transcripts per million [TPM] ~ 25) (22, 23). CAR is the high-affinity receptor ($K_D \sim 0.2 \mu\text{M}$) that is less abundant overall (median TPM ~ 3.5) and restricted to the heart, brain, and epithelial tissues that CVB3 infects (22, 24). DAF is a GPI-linked surface protein that diffuses freely in cholesterol-rich microdomains on the cell surface, whereas CAR is a cell-cell adhesion protein localized to tight junctions. Binding of CVB3 to DAF enables the CVB3:DAF complex to traffic to tight junctions, where CVB3 is bound by CAR to promote endosomal internalization (Figure 2.3A) (25). In the model, we assumed that CAR is instantly degraded upon CVB3 internalization at characteristic rates for endocytosis (26). However, results in permissive hosts were unchanged if CAR was instead recycled instantaneously to the surface (Figure 2.4A). Rates of rapid endosomal escape were estimated from live-cell experiments with labeled poliovirus (19). The output of the delivery module is positive-strand RNA (both infectious and defective) in the cytoplasm.

As a representative host cell, we used AC16 cells, which were originally immortalized by cell fusion of adult ventricular cardiomyocytes with SV40-transformed fibroblasts (27). AC16 cells are not very permissive to CVB3 infection, but they become highly susceptible upon ectopic expression of CAR (28). We originally intended to use CAR-overexpressing AC16 cells (AC16-CAR) under quiescent conditions in which SV40 was knocked down and serum reduced (27). However, RNA sequencing (RNA-seq) revealed that CAR overexpression in quiescent cells caused a considerable upregulation of ISGs (Figure 2.3B), which would confound the analysis. Follow-up experiments revealed that ISG upregulation was specific to the quiescence protocol (Figures 2.3C and 2.4B–D), enabling the use of proliferating AC16-CAR cells for the study.

The delivery module is composed entirely of low copy-number events and intermediate species that are difficult to assess experimentally. Consequently, we selected DAF and CAR abundance for a validity check on the module, using the other modules to amplify the effect of CVB3 receptors on viral outputs that were measurable. To parameterize initial conditions for DAF and CAR, we derived estimates from the AC16 RNA-seq data by using an RNA-to-protein conversion factor for each receptor estimated from a panel of different HeLa lines profiled by transcriptomics and proteomics (29, 30). HeLa cells are highly susceptible to CVB3 and are widely used for virus propagation; thus, per-cell receptor abundances should be in the operating range for productive infection. Both

conversions were adequately described by a hyperbolic-to-linear fit providing approximate protein estimates for parental AC16 and AC16-CAR cells (Figures 2.3D and 2.3E; Section 2.5, STAR Methods). However, CAR estimates from corresponding *CXADR* mRNA in AC16-CAR cells required an enormous extrapolation of the fit (Figure 2.3E), prompting an independent measurement. Ectopic CAR contains a V5 tag on the C-terminus, which prevents the construct from transmitting adhesion signals to PDZ domain-containing proteins (31) and enables absolute epitope quantification. Using a recombinant multitag protein containing a V5 epitope, we calibrated V5 immunoblotting to protein copies and interpolated CAR-V5 copies for known numbers of AC16-CAR cells (Figures 2.3F and 2.4E). The ~5.5 million CAR-V5 molecules per cell falls within the prediction interval for the RNA-to-protein conversion fits, supporting that fit-based estimates for DAF (~61,000 [AC16] and ~44,000 [AC16-CAR] per cell) and endogenous CAR (~2300 per cell) are realistic.

Taking an average DAF abundance between the two lines, we simulated the quantitative importance of CAR at endogenous and ectopic levels with a CVB3 infection of 10 PFU. When viral particles were fully accounted for, the model predicted a robust infection in CAR-overexpressing cells but not parental AC16 cells, reflected by viral protein 1 (VP1) expression within 10 hours (Figures 2.1D and 2.3G). Host permissiveness was predicted to occur at ~7000 CAR copies per cell (Figure 2.4F), a threshold surpassed by all HeLa variants in the panel (Figure 2.3E) (30). By contrast, when defective particles were ignored and the same host-cell conditions simulated, productive CVB3 infections occurred at both ectopic and endogenous abundances of CAR (Figure 2.3H). Lack of receptor competition from defective particles enables DAF-bound PFUs to access the small number of CARs, internalize, and replicate. Competition was consequential at particle-to-PFU ratios of 200 or greater (Figure 2.4G), partially overlapping with the range of ratios documented for enteroviruses (2). The results suggest that host-cell permissiveness is conditioned on PFU purity relative to defective interfering particles.

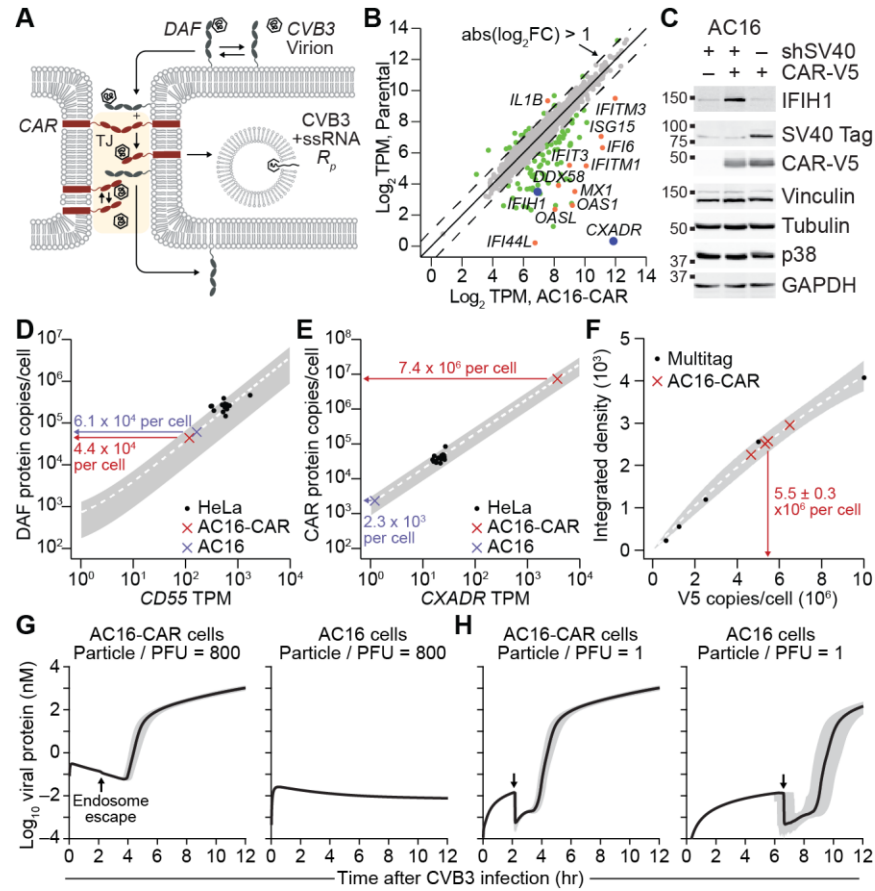


Figure 2.3. Stoichiometric estimation and simulation of CVB3 particles, cargo, and surface receptors.

(A) Overview of the CVB3 delivery module. CVB3 binds to DAF and translocates to tight junctions (TJ, beige). In the tight junction, CVB3 unbinds DAF and binds to CAR by two parallel mechanisms (Section 2.5, STAR Methods). CAR-bound CVB3 is internalized and DAF is returned to the plasma membrane. At any point, CVB3 may dissociate from its receptors (reverse arrows). After internalization, the viral genome (R_p) escapes the endosome into the cytoplasm.

(B) Ectopic expression of CAR in quiescent AC16 cells causes chronic overexpression of immune response genes. Differentially expressed transcripts between quiescent parental AC16 cells (Parental) and CAR-overexpressing AC16 cells (AC16-CAR) ($q < 0.05$; gray) that also have a \log_2 fold change (FC) greater than ± 1 (green) and are immune-regulated genes (orange). Proteins encoded by *IFIH1* and *CXADR* (blue) are independently measured in (C).

(C) The elevated interferon response of AC16-CAR cells is restricted to the quiescence protocol involving SV40 knockdown. AC16 cells with or without quiescence (shSV40) or CAR overexpression (CAR-V5) were immunoblotted for IFIH1, SV40 T antigen (Tag), and ectopic CAR-V5 with vinculin, tubulin, p38, and GAPDH used as loading controls. Replicated densitometry is shown in Figures 2.4B–D.

(D and E) Estimating DAF–CAR abundance through coupled transcriptomics and proteomics of HeLa cells (30). Data from $n = 14$ HeLa variants was used to determine a hyperbolic-to-linear fit (Section 2.5, STAR Methods) that was combined with the AC16 transcriptomic data to estimate the per-cell DAF–CAR protein abundances for AC16 cells (blue) and AC16-CAR cells (red). The best-fit curve (white dashed) is shown $\pm 99\%$ confidence interval of the fit (gray).

(F) Direct measurement of CAR abundance in AC16-CAR cells. Recombinant V5-containing Multitag was used to calibrate a V5 quantitative immunoblot and estimate the total per-cell abundance of CAR-V5. Data are shown as the mean \pm s.e.m. of $n = 4$ biological replicates (red). The best-fit calibration curve (white dashed)

is shown \pm 90% confidence interval of the fit (gray).

(G and H) Complete kinetics requires accounting of defective CVB3 particles to capture host-cell tropism. For these simulations, DAF abundance was averaged to 52,500 copies per cell. Predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%. In **(H)**, the lack of defective particles exaggerates the drop in viral protein associated with endosomal escape (arrows; Section 2.5, STAR Methods). Results from **(G)** and **(H)** were unchanged if endogenous CAR was estimated from a calibration including the protein estimates from AC16-CAR cells in **(F)**.

See also Figure 2.4.

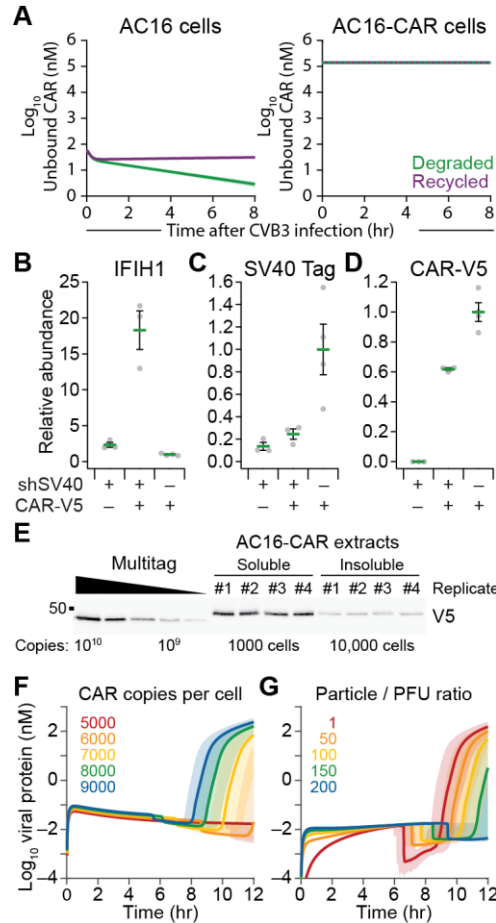


Figure 2.4. Delivery controls, immunoblot quantification, V5 immunoblot image, and evaluation of receptor-ratio thresholds.

(A) Recycling of CAR upon CVB3 internalization is inconsequential for permissive hosts. Nonpermissive (AC16, left) and permissive (AC16-CAR, right) complete kinetic models are simulated assuming total degradation (green) or recycling (purple) of CAR upon CVB3 internalization. AC16 cells are nonpermissive under both assumptions. Predictions of unbound CAR are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

(B–D) Replicated densitometry of IFIH1, SV40 Tag, and CAR-V5. Data are shown as the mean \pm s.e.m. from $n = 3$ –4 biological replicates.

(E) V5 immunoblot image of V5-containing Multitag recombinant protein along with RIPA-soluble (left) and RIPA-insoluble (right) extracts of AC16-CAR cells at the indicated cell abundance per lane. The abundance of soluble and insoluble CAR-V5 was combined per biological replicate in Figure 2.3F.

(F) Minimum CAR abundance for productive infection of AC16 cells with CVB3. Predictions of total viral protein for 5000–9000 CAR molecules per cell are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

(G) Critical particle-to-PFU ratio for accurately predicting the lack of CVB3 tropism in AC16 cells. Predictions of total viral protein for particle-to-PFU ratios from 1–200 are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

2.3.3 A two-phase encoding of membrane replication predicts quantifiable positive–negative strands and replicative intermediates

After entering the cytoplasm, an infectious positive-strand RNA genome is translated by host-cell ribosomes. In cell-free systems, enteroviral RNA is recognized by 2–5 ribosomes within 15 seconds (32). We conservatively assumed a polysome size of 2–3 ribosomes to account for ribosome competition with host-cell mRNAs early in infection. In the model, polysomes increase the net translation of viral proteins along with the retention of positive strands in translation complexes (Section 2.5, STAR Methods). Modeling polysomes was critical for appropriately timing the onset of exponential viral protein synthesis under realistic translation rates (Figure 2.6A).

Any positive strand released from a translation complex was instantly placed on a 3D^{pol}-containing surface to capture the cis coupling of enteroviral translation and replication (Figure 2.5A; Section 2.5, STAR Methods) (33, 34). During infection, host-cell membranes are redirected to create molecular factories for genome replication from a negative-strand template (35). The formation of such “viral replication organelles” (VROs) is widely documented in positive-strand RNA viruses and may serve multiple functions (36). Using the model, we considered three functions for VROs: 1) shield viral RNAs from degradation, 2) shield viral replicative intermediates from dsRNA sensing, and 3) accelerate replicative processes by concentrating species on a surface. Recent ultrastructural studies suggest a per-cell VRO surface area of 160–185 μm^2 (37), which roughly agrees with our own brightfield estimates of 120 μm^2 (Section 2.5, STAR Methods). These numbers and the ~7-nm height of a 3D^{pol} enzyme (PDB ID: 3CDW) yield a >2000-fold concentrating effect of associating cytoplasmic viral molecules in the local volume of a VRO surface. Whereas setting either viral RNA degradation or dsRNA sensing on VROs to zero had a negligible impact on the timing of acute infection, reducing the VRO-concentrating effect to less than 100-fold yielded no net output (Figure 2.5B). Other VRO functions might be important for RNA viruses that replicate more slowly and organize VROs differently (38), but these computational results strongly suggest that the dominant role for enteroviral VROs is to accelerate biochemistry.

Next, we needed to define when the VRO acceleration was triggered kinetically. One 3D^{pol} enzyme by itself is dilute whether in the cytoplasm or on a membrane, but local surface patches of enzyme and positive strand could exhibit acceleration even though the whole-cell concentration is low. Guided by measurements of positive and negative strands in CVB3-infected cells (Section 2.5, STAR Methods), we selected a threshold of 25 3D^{pol} molecules, which triggers at ~2.5 hr after infection with 10 PFU and coincides with the first

translational burst of an infectious virion (39). This threshold marks the early onset of the deterministic, continuous regime ($3D^{\text{pol}}$ counting noise = 20%) and precedes the 4–5-hr time point when VROs become ultrastructurally observable (40).

To validate the parameterization of the replication module, we sought robust, sensitive, and absolute measurements of positive and negative strands during infection. We devised a tagged quantitative PCR (qPCR) assay, which avoids false priming within the CVB3 genome by using a biotinylated strand-specific primer and streptavidin pulldown before quantification (Section 2.5, STAR Methods) (41, 42). Using purified standards, the assay was sensitive to ~1000 CVB3 copies per 250-cell reaction and linear over at least five decades (Figures 2.6B and 2.6C). After specifying a VRO transition consistent with these data, the model achieved excellent quantitative agreement with the absolute estimates of positive–negative strands (Figures 2.5C and 2.6D), and the module was deemed valid. Results were comparable if defective RNA genomes assembled as polysomes in the model or were further subject to nonsense-mediated decay (Figures 2.6E and 2.6F). Future iterations will consider the *de novo* generation of defective genomes during replication (43).

Model–experiment concordance required the particle-to-PFU ratio described earlier and a quantification of the positive-to-negative strand ratio in released particles (1790; Figure 2.6B). Notably, we also assumed that replicative intermediates—partial RNA–RNA hybrids of positive and negative strand—were not detected by the tagged qPCR method. Upon cell lysis, collapsed replicative intermediates would be difficult to denature fully for biotinylated priming without thermally degrading the RNA itself. We tested the assumption by mixing 10^5 copies of purified strand with increasing amounts of the complementary strand per 250-cell reaction and measuring both with the stranded assay (Section 2.5, STAR Methods). Although copy numbers were accurate for many mixtures, some limiting ratios were irreproducible or grossly underestimated (Figure 2.5D). Such strand competition is probably even more severe in cells, where replicating strands are already nearby their complementary template (44).

There are experimental workarounds to accessing collapsed replicative intermediates, but their detection efficiency is difficult to determine (44–46). Using the model at 10 PFU, we simulated the positive-to-negative strand ratio predicted when replicative intermediates were incorporated in the calculation at different efficiencies (Section 2.5, STAR Methods). Replicative intermediates, even at 10% detection efficiency, profoundly altered the calculated ratio and its dynamic trajectory (Figure 2.5E). Literature-derived ratios of 50:1 to 100:1 were recapitulated in the model with detection efficiencies of 10–30% (44, 45, 47). At 100% detection efficiency, the model predicted lower ratios near

15:1. Lower positive-to-negative strand ratios have been observed in single-cell enterovirus assays involving large dilutions of a cell extract after lysis that may disfavor collapse of replicative intermediates (48). Taken together, we conclude that the replication module is consistent with internal measurements as well as a range of observations in the literature.

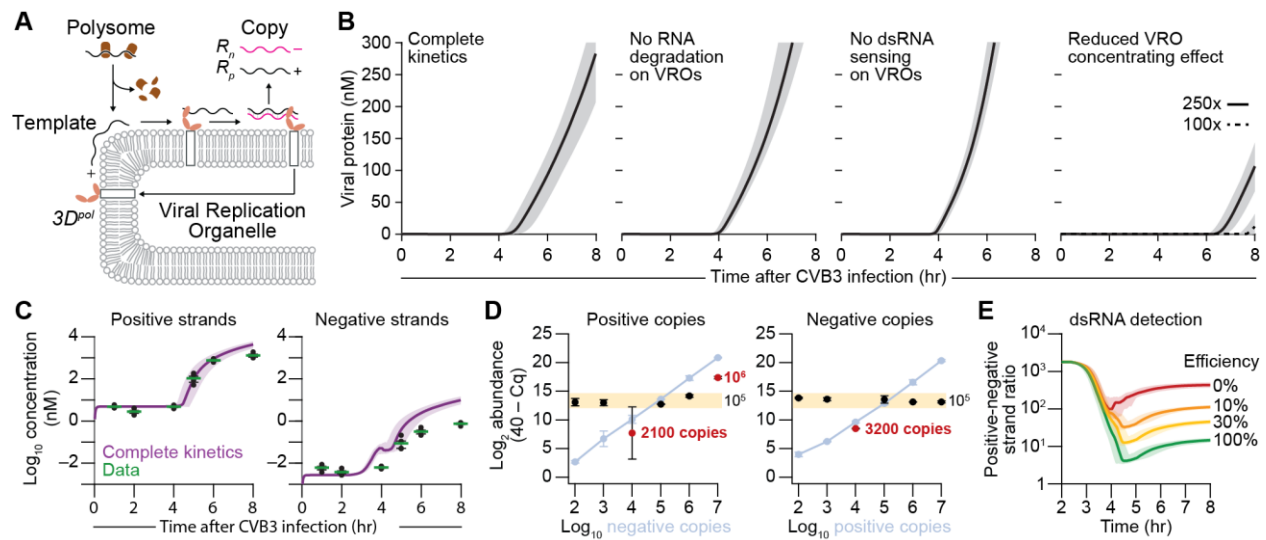


Figure 2.5. Explosive genome replication requires intracellular membranes and hides negative-strand templates from quantitation.

(A) Overview of the replication module. After dissociating from polysomes, the positive-strand RNA genome is released to associate with the surface of viral replication organelles (VROs), where translated RNA-dependent RNA polymerase ($3D^{pol}$) resides bound to 3AB (white box). Replication proceeds on the VRO surface, releasing one positive strand (R_p) and one negative strand (R_n). The reciprocal process occurs with R_n as template, except that R_n is not released from $3D^{pol}$ (Section 2.5, STAR Methods).

(B) VROs are surface accelerants. Translational output from complete kinetics altered to assume zero RNA degradation, zero dsRNA sensing, or limited concentrating effect on the VRO surface. The concentrating effect in the complete kinetic model is 3216x (see Section 2.5, STAR Methods).

(C) Complete kinetics captures the absolute viral RNA dynamics of positive and negative strands. Predictions (purple) were compared to data (green) obtained by strand-specific tagged quantitative PCR (qPCR) with purified standards. Data are shown as the geometric mean \pm log-transformed standard error of $n = 4$ biological replicates of AC16-CAR cells infected at MOI = 10 for the indicated times. Population-level simulations are shown in Figure 2.6D.

(D) Strand competition between sense and antisense CVB3 genomes in vitro. Strand-specific tagged qPCR of 10^5 positive copies (left) or negative copies (right) amidst the indicated abundance of complementary strand on the x-axis (measured in blue). Data [black (with outliers highlighted in red) and blue] are shown as the mean \log_2 relative abundance $[40 - \text{qPCR quantification cycle (Cq)}] (42) \pm$ range of assay duplicates at $n = 6$ separate positive-negative strand mixtures.

(E) Observed positive-negative strand ratio depends critically on the detection efficiency of dsRNA replicative intermediates. The complete kinetic model was simulated and inventoried with different fractional contributions of dsRNA to the positive- and negative-strand totals.

For (B), (C), and (E), predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

See also Figure 2.6.

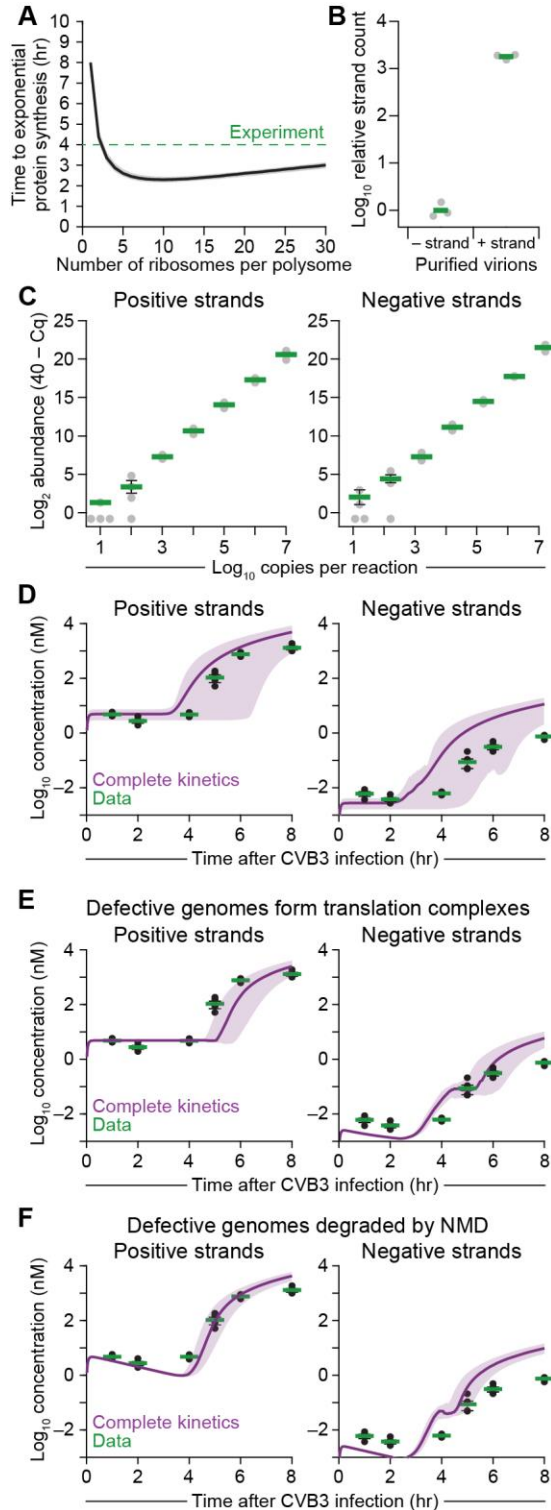


Figure 2.6. Polysome size requirements, strand-specific tagged qPCR validation, and population-level simulations of viral RNA strand dynamics with complete kinetics.

(A) Polysome size dictates the timing of viral protein synthesis. The start of exponential protein synthesis is shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%. The experimentally observed timing

(Figure 2.1D) is shown for reference (green).

(B) Strand-specific tagged qPCR assessment of positive (+) and negative (–) strands in purified CVB3 virions. The measured strand ratio was 1790 ± 130 . Data are shown as the geometric mean of $n = 3$ separate CVB3 dilutions of purified virion.

(C) Quantitative precision and accuracy of the strand-specific tagged qPCR assay with positive strands from purified CVB3 virions and negative strands prepared by in vitro transcription. Data are shown as the mean \log_2 relative abundance $[40 - \text{qPCR quantification cycle (Cq)}]$ $(42) \pm \text{s.e.m.}$ of assay quadruplicates at $n = 7$ separate positive-negative strand dilutions. Note the sporadic detection of the assay as the input approaches the single-molecule limit (assuming $\sim 10\%$ conversion efficiency from RNA to amplifiable cDNA).

(D) Complete kinetics captures the absolute viral RNA dynamics of positive and negative strands at the population level.

(E and F) Alternative fates of defective CVB3 genomes do not substantially affect the modeled viral RNA dynamics of positive and negative strands. Defective genomes were assumed to **(E)** form defective polysome translation complexes that bound–unbound but did not yield polyprotein or **(F)** form defective polysome translation complexes that bound and instantly underwent nonsense-mediated decay (NMD).

For **(D–F)**, predictions (purple) were compared to data (green) obtained by strand-specific tagged quantitative PCR (qPCR) with purified standards. Data are reprinted from Figure 2.5C and shown as the geometric mean \pm log-transformed standard error of $n = 4$ biological replicates of AC16-CAR cells infected at MOI = 10 for the indicated times. Predictions are shown as the mean simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of population-level infections at MOI = 10 **(D)** or 100 simulations of single-cell infections at 10 PFU **(E and F)** with a parameter coefficient of variation of 5%.

2.3.4 Encapsidation must coordinate the kinetics of enteroviral protein synthesis, recruitment, and self-assembly

Besides 3D^{pol}, positive-strand translation also yields equal numbers of enteroviral proteinases, hydrophobic membrane-interacting proteins, and structural proteins that form the viral capsid (Figure 2.1A). We simplified by defining a single, lumped protease with the substrate specificity of both 2A^{pro} and 3C^{pro} (Section 2.5, STAR Methods). Likewise, membrane-associated 3D^{pol} in the model considers 3D^{pol} together with the tightly interacting hydrophobic protein, 3AB (Figure 2.5A) (49). We grouped the remaining hydrophobic proteins into a single exemplar, 2C^{ATPase}, because of its direct interaction with the VP3 structural protein (50). VP3 assembles instantaneously with VP0 and VP1 to form a VP0–VP1–VP3 protomer (51). We balanced molecular detail and complexity by modeling structural proteins as the 14S intermediate pentamer [(VP0–VP1–VP3)₅], one fifth of which was generated with each polyprotein matured (Figure 2.7A). Together, these simplifications decompose the mature CVB3 polypeptide into four species—protease, 3D^{pol}, 2C^{ATPase}, and pentamer—that should balance stoichiometrically under limiting conditions (Figures 2.8A and 2.8B).

One of the earliest consequences of enteroviral proteinase maturation is the cleavage of the host-cell eukaryotic initiation factor, eIF4G, by 2A^{pro} (17). eIF4G cleavage prevents ribosomes from initiating cap-dependent translation, thereby favoring the cap-independent translation of the virus (Figure 2.7A) (52, 53). We modeled the shutoff of host-cell translation and theft of ribosomes by incorporating a protease-catalyzed conversion of ribosomes from inaccessible (cap-dependent) to accessible for CVB3 (Section 2.5, STAR Methods). The full model showed good agreement with the relative dynamics of eIF4G cleavage and the synthesis of VP1 capsid protein (Figures 2.7B, 2.7C, 2.8C, and 2.8D), suggesting accurate encoding of the precursors to encapsidation.

Viral pentamers were not placed immediately on VROs like 3D^{pol} but instead were recruited to membranes by their interaction with available 2C^{ATPase} (Figure 2.7A; Section 2.5, STAR Methods) (50). As neither kinetic nor equilibrium constants are available for the VP3–2C^{ATPase} interaction, we began with rate parameters that were realistic and consistent with a nominal affinity of 100 nM (Table S1). For viral RNA of either strand, there are multiple possible binding partners (54), and thus we modeled VRO recruitment as a passive exchange between cytoplasmic and VRO compartments (Section 2.5, STAR Methods). Considering the VRO concentrating effect and assuming a representative RNA-protein association rate of 25 nM⁻¹hr⁻¹ (55), an exchange rate of 1 hr⁻¹ equates to an effective membrane affinity of ~125 nM. Sensitivity of the model to these approximations would be

analyzed after defining the encapsidation steps that yield mature virions.

Encapsidation of enteroviral RNA arises from a series of individually weak interactions (~ 1 mM contact affinity) that multiplicatively contribute to the final “closed” virion (56), which we assumed to be irreversible (Figure 2.7D and Table S1; Section 2.5, STAR Methods). Capsid self-assembly can also take place without viral RNA. We simplified the combinatorics by assuming that assembly occurs one pentamer at a time through reversible additions of pentamers that are either unbound or bound to RNA (Section 2.5, STAR Methods). In this formalism, intermediate pentamer states can arise from multiple paths. For example, an empty state of five pentamers (*P5Empty*) can result from the addition of a pentamer to *P4Empty*, the loss of a pentamer from *P6Empty*, or the loss of RNA-bound pentamer from *P6Filled*. The disassembly of filled intermediates was assumed to release RNA with decreasing probability as the number of pentamers in the intermediate increased. *P6Filled*, for instance, disassembles to *P5Filled* + free pentamer $\frac{5}{6}$ of the time and *P5Empty* + RNA-containing pentamer $\frac{1}{6}$ of the time. The detailed encoding of encapsidation enables the module to keep track of the discrete steps to virion production and identify stoichiometric depletions of precursors or pentamer states if they occur.

To illustrate the importance of weak and balanced interactions for encapsidation, we increased the contact affinity between RNA and pentamer from 1 mM to 100 μ M and found that the slight imbalance completely blocked encapsidation (Figure 2.7E). Pentamers recruited to the VRO were fully sequestered by the increasing abundance of positive strands, depleting the free pentamer pool available for higher-order assembly of capsids. Similar kinetic-trapping mechanisms dominated when pentamer contact affinities were tighter than 10 μ M. By comparison, the module was not as sensitive to the VP3–2C^{ATPase} interaction strength that recruits pentamer to VROs, with virtually no change observed when the affinity was altered tenfold (Figure 2.7F and Table S1). Unrealistic delays in encapsidation did not occur until the affinity was reduced from 100 nM to 1–10 mM, suggesting that alanine mutants of 2C^{ATPase} disrupting encapsidation must severely hinder interactions with pentamer (57).

In contrast to pentamer recruitment, we found that kinetic interactions between viral RNA and the VRO surface were critical. Increasing the effective membrane affinity (tenfold decrease in k_{off} and thus K_D ; Figure 2.7A) largely blocked capsid formation by driving RNA prematurely to the VRO at the expense of translation complexes needed to generate sufficient 3D^{pol} for replication (Figures 2.7G–I). Even with the same equilibrium affinity, faster exchange was problematic for effective encapsidation. We noted virion production tailed away when on–off rates were both increased tenfold (Figure 2.7J). The shorter

residence time at the VRO led to large increases in early pentamer species (*P1Filled* to *P6Filled*) at the expense of later filled pentamers needed for sustained virion production (*P9Filled* to *P11Filled*; Figure 2.7K). The model thus provides a quantitative rationale for the multiple RNA-binding interactions at the membrane surface (51). Overall, the encapsidation module finishes the viral life cycle in a stoichiometrically consistent way and makes specific predictions about the critical kinetic and thermodynamic steps of self-assembly.

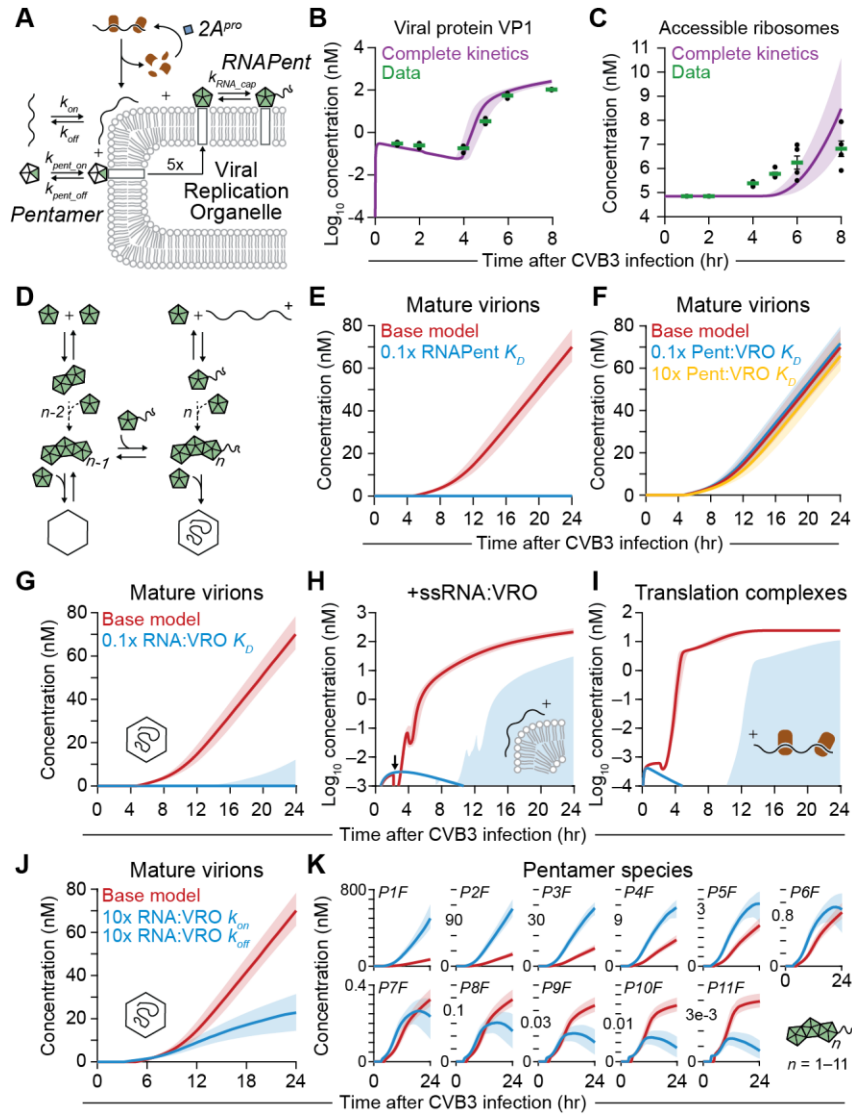


Figure 2.7. A Goldilocks zone for enteroviral encapsidation.

(A) Overview schematic of the CVB3 encapsidation module on the viral replication organelle (VRO). After viral protein translation (aided by $2A^{pro}$ cleavage of eIF4G), viral capsid pentamers are recruited to VROs by binding $2C^{ATPase}$ (white box). At the VRO, pentamers associate with one another and with positive-strand genomes to form RNA-pentamer assemblies (RNA_{Pent}). Rate parameters investigated by perturbation are shown.

(B and C) Complete kinetics captures the dynamics of viral protein VP1 expression and accessible ribosomes (estimated from viral protease cleavage of host eIF4G). Predictions (purple) were compared to data (green) obtained by quantitative immunoblotting. Data are shown as the geometric mean \pm log-transformed standard error of $n = 4$ biological replicates of AC16-CAR cells infected at MOI = 10 for the indicated times. Population-level simulations are shown in Figures 2.8C and 2.8D.

(D) Overview schematic of the CVB3 encapsidation module showing viral capsid assembly. Capsids assemble one pentamer at a time with or without a positive-strand genome. Filled virions (lower right) are formed irreversibly.

(E) Increasing RNA-Pentamer affinity prevents virion maturation. Virion production in the base model (red) compared to when the RNA-pentamer affinity is increased (blue) by reducing its contact dissociation constant ($RNA_{Pent} K_D$) from 1 mM to 100 μ M (Section 2.5, STAR Methods).

(F) Model simulations are not sensitive to changes in pentamer recruitment to the VRO via $2C^{ATPase}$ binding. Virion production in the base model (red) compared to when pentamer- $2C^{ATPase}$ affinity is either increased (blue) or decreased (yellow) by reducing or increasing its apparent dissociation constant (Pent:VRO K_D) from 100 nM to 10 nM (blue) or 1 μ M (yellow).

(G–I) RNA interactions with the VRO must be balanced. (G) Virion production, (H) positive-strand RNA genomes at the VRO (+ssRNA:VRO), and (I) translation complexes in the base model (red) compared to when RNA exchange rates with the VRO were biased (blue) by decreasing k_{off} to from 1 hr^{-1} to 0.1 hr^{-1} . Arrow in (H) indicates premature recruitment of positive-strand RNA genomes to the VRO when $k_{off} = 0.1 \text{ } hr^{-1}$. Some lognormally sampled parameter sets give rise to the late formation of +ssRNA:VRO, productive translation complexes, and mature virions (blue shading).

(J and K) Kinetics of RNA-VRO interactions are important for sustained virion production. (J) Virion production and (K) pentamer states in the base model (red) compared to when RNA association and dissociation rates with the VRO are both increased tenfold (blue). PnF, intermediate filled capsid state of n pentamers and one positive-strand genome.

For (B), (C), and (E–K), predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

See also Figure 2.8.

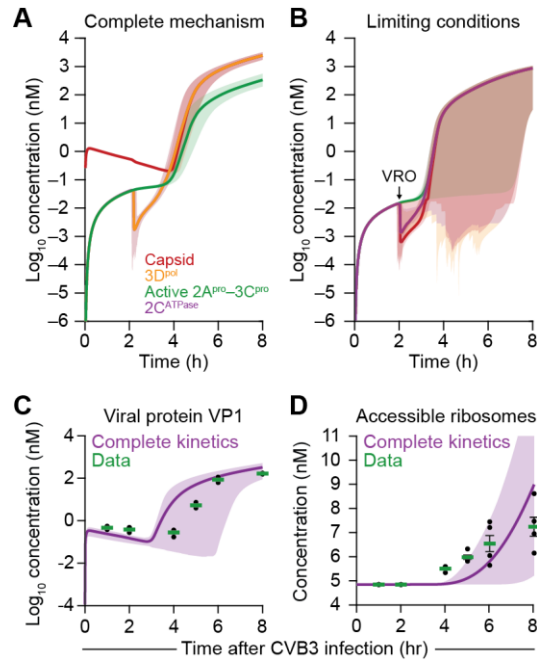


Figure 2.8. Stoichiometric balance of CVB3 protein classes and population-level simulations of viral-protein and eIF4G cleavage dynamics.

(A) Dynamics for the four protein species—capsid protein, $3D^{pol}$, active protease ($2A^{pro}$ – $3C^{pro}$), and $2C^{ATPase}$ —in the complete kinetic model.

(B) Dynamics for the four protein species under the following limiting conditions for stoichiometric balance: no protein-RNA degradation, no antiviral response, no viral antagonism, no defective particles, and no bookkeeping of viral capsid from added PFUs. The separation of the four species at ~two hours is caused by the immediate onset of the viral replication organelle (VRO) regime, which affects the apparent concentration of all species except active protease (Section 2.5, STAR Methods). The solutions converge again within two hours.

(C and D) Complete kinetics captures the dynamics of viral protein VP1 expression and accessible ribosomes (estimated from viral protease cleavage of host eIF4G) at the population level. Predictions (purple) were compared to data (green) obtained by quantitative immunoblotting. Data are reprinted from Figures 2.7B and 2.7C and shown as the geometric mean \pm log-transformed standard error of $n = 4$ biological replicates of AC16-CAR cells infected at MOI = 10 for the indicated times. Predictions are shown as the mean simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of population-level infections at MOI = 10 with a parameter coefficient of variation of 5%.

For (A) and (B), predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

2.3.5 Loss of type I interferon signaling coincides with degradation of MAVS during CVB3 infection

With the viral life-cycle modules encoded, we returned to host-cell signaling responses in search of pathways engaged during an acute CVB3 infection. The CVB3 proteinase 3C^{pro} can block inflammatory signaling by cleaving the NF- κ B inhibitor, I κ B α , and converting it to a stably-associated, non-degradable form (58). In AC16-CAR cells, however, we detected canonical degradation of I κ B α and induction of NF- κ B target genes a few hours after CVB3 infection (Figures 2.9A and 2.9B). We next considered the type I interferon signaling pathway, which is robustly triggered by cytosolic dsCVB3 in these cells (Figure 2.1B). Although phosphorylation of the effector kinase TBK1 was detected transiently after infection, the antiviral signal was not reliably propagated to IRF3 or interferon signaling through phosphorylated STATs (Figures 2.9C and 2.9D). There was no induction of the *IFNA* locus, nor did we observe upregulation of any ISG profiled in cells infected with live virus (Figures 2.9E and 2.10A–E). The results suggested that CVB3 actively severs signal transmission from its dsRNA replicative intermediate to a productive type I interferon response.

Both 3C^{pro} and 2A^{pro} are reported to target components of the innate dsRNA sensing and signal-transduction machinery (5, 15). Normally, long dsRNA is detected by MDA5, which associates with the mitochondrial transducer MAVS to drive surface polymerization that signals through the adaptor TRAF, the kinase TBK1, and the transcription factor IRF3 (59). Although MDA5 was not significantly altered in AC16-CAR cells during CVB3 infection, we noted a mid-infection dip in proteoforms of the mitochondrial transducer, MAVS (Figure 2.9F). Later post-infection times revealed a C-terminal fragment at 35 kDa (MAVS₃₅), which comprised up to 20% of the endogenous MAVS protein at 24 hours; MDA5 was not clearly affected over the same timeframe (Figures 2.9G and 2.9H). The mid-infection decrease in full-length MAVS (MAVS_{FL}) was concurrent with the loss of TBK1 phosphorylation and the explosive increase in CVB3 protein (Figures 2.7B, 2.9C, and 2.9F). A MAVS₃₅ fragment implied that MAVS_{FL} was split in half, separating its N-terminal oligomerization domain from the C-terminal mitochondrial–microsomal transmembrane domain (60, 61). Even slight abundance shifts in MAVS proteoforms can alter the propensity of the pathway to activate (62, 63). Together, the data suggested that a CVB3-derived protein antagonizes the dsRNA-mediated interferon response by cleaving an essential transducer (64).

We elaborated the infection model with a lumped interferon response, which pairs a CVB3 sensing-and-transduction mechanism with a single, pleiotropic ISG effector (*ISGs*,

Figure 2.1A; Section 2.5, STAR Methods). The joint sensor–transducer recognizes dsRNA and sigmoidally induces the ISG effector, streamlining the native regulation by the MDA5–MAVS–TRAF–TBK1–IRF3 pathway (65). The ISG effector impedes CVB3 infection at three points in the life cycle. First, the effector accelerates CVB3 RNA turnover, modeling the activity of oligoadenylate synthetases and RNase L (Figures 2.1A and 2.1B) (66). Second, the effector blocks CVB3 polyprotein synthesis to capture dsRNA recognition and translational inhibition by the PKR pathway (Section 2.5, STAR Methods) (67). Third, the effector incorporates interferon-stimulated mechanisms for disabling CVB3 proteinases, such as ISGylation of 2A^{pro} and PARP9–DTX3L-mediated ubiquitylation of 3C^{pro} (68, 69). The three antiviral effects were encoded as hyperbolic feedbacks on the CVB3 RNA degradation rates, the formation rate of CVB3 translation complexes, and the effective production rate of active 2A^{pro}–3C^{pro} respectively (Figure 2.1A; Section 2.5, STAR Methods). The active 2A^{pro}–3C^{pro} pool reciprocally feeds back on the sensor–transducer in the model to hyperbolically limit the maximum induction rate of the ISG effector in response to dsRNA (Figure 2.1A; Section 2.5, STAR Methods). The interlinking of these four negative feedbacks provides a compact abstraction of the multifaceted antagonism between enteroviruses and type I interferon signaling.

The feedback parameters of the lumped interferon response are phenomenological and not directly measurable. Nevertheless, the feedback architecture places strong qualitative constraints on their relative potencies:

1. AC16-CAR cells are capable of mounting an interferon response to CVB3 dsRNA (Figure 2.1B). However, the response is blocked during an active infection, presumably through protease-mediated antagonism of dsRNA sensing and cleavage of MAVS (Figures 2.9E–H). The feedback potency of the 2A^{pro}–3C^{pro} pool must therefore be potent enough to degrade dsRNA sensing in the model during an active infection.
2. Early stimulation of AC16-CAR cells with interferons significantly impedes CVB3 propagation (28), suggesting that an endogenous interferon response would be sufficient if it were fully mobilized and not disrupted by viral proteases. Thus, the modeled type I interferon response should thwart the virus in simulations where dsRNA sensing–transduction is fully intact.
3. At later times during CVB3 infection, the addition of interferons becomes ineffective (Figure 2.10F). The 6–8 hr time window indicates that antiviral mechanisms become ineffective after the virus has passed a critical “point of no return” in its life cycle.

Using these constraints, we identified feedback weights with half-maximal effective concentrations (EC50s) that conferred the specified qualitative behavior (Figure 2.9I; Section 2.5, STAR Methods). The outcome of infection was robust when these heuristic EC50 values were varied individually over a severalfold range (Figure 2.10G). Most EC50 values were in the nM range (5–20 nM) except for protease antagonism of dsRNA sensing–transduction, which needed to be much lower (1 pM). Such potency could be achieved if the relevant protease is locally complexed with the dsRNA sensor–transducer that is cleaved. For MAVS cleavage, a plausible candidate is the precursor to 3C^{pro} and 3D^{pol}, 3CD^{pro}, which has both protease activity and nucleotide-binding activity for replication (70). The importance of potentially antagonizing ISG induction is corroborated by a recent computational model of Dengue virus (71). Combining the enteroviral life cycle with host-cell feedbacks created a rich dynamical system to mine for predicted behavior that was non-intuitive and possibly important for pathogenesis.

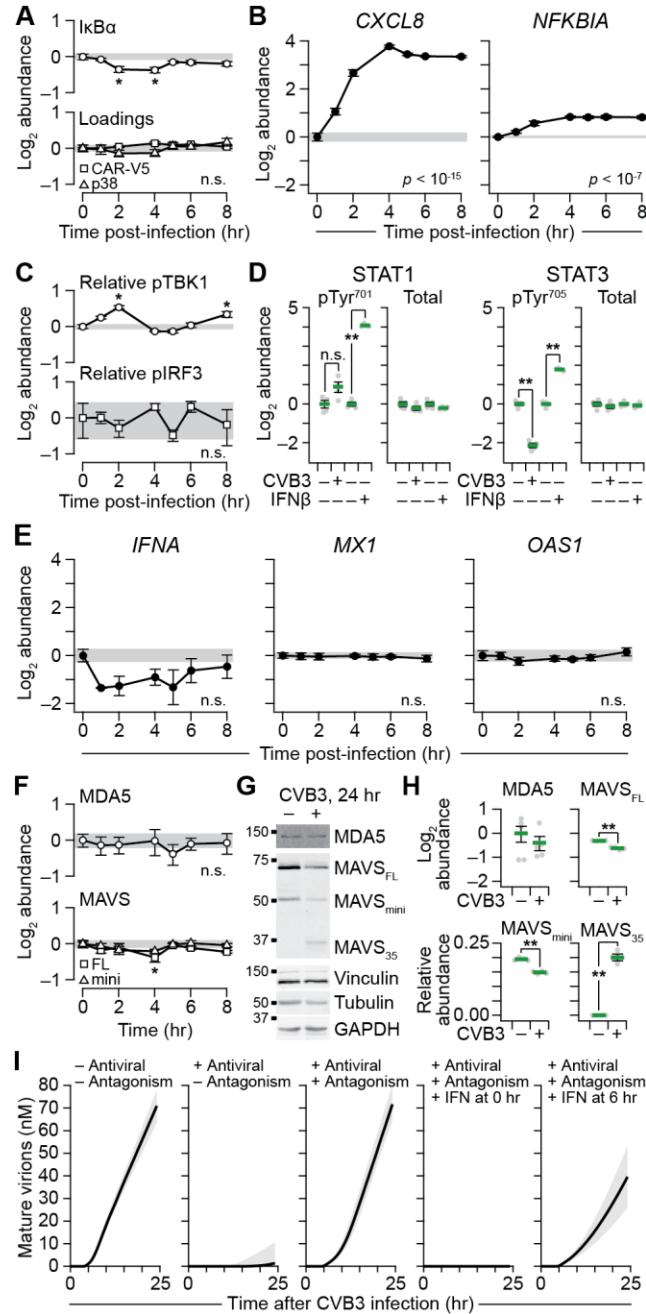


Figure 2.9. CVB3 partially dismantles antiviral signaling and the dsRNA transducer MAVS.

(A and B) Acute CVB3 infection activates NF-κB signaling. (A) AC16-CAR cells were infected with CVB3 at MOI = 10 for the indicated times and immunoblotted for IκBα (A, upper) with ectopic CAR-V5, p38 (A, lower), vinculin, and tubulin used as loading controls. (B) Total RNA was collected at the indicated times and measured for the indicated NF-κB target genes with *HINT1*, *PRDX6*, and *GUSB* used as loading controls.

(C–E) Early dsRNA sensing is not propagated to a detectable type I interferon response during acute CVB3 infection. (C) AC16-CAR cells were infected with CVB3 at MOI = 10 for the indicated times and immunoblotted for phosphorylated TBK1 (Ser¹⁷²) relative to total TBK1 (Relative pTBK1) (C, upper) or phosphorylated IRF3 (Ser³⁹⁶) relative to total IRF3 (Relative pIRF3) (C, lower) with ectopic CAR-V5, p38, vinculin, and tubulin used

as loading controls. **(D)** AC16-CAR cells were infected with CVB3 at MOI = 10 for 24 hours or treated with 30 ng/ml IFN β for 30 minutes as a positive control and immunoblotted for phosphorylated STAT1 (pTyr⁷⁰¹) and total STAT1 (**D**, left) or phosphorylated STAT3 (pTyr⁷⁰⁵) and total STAT3 (**D**, right) with vinculin, tubulin, and GAPDH (\pm CVB3) or actin, tubulin, and p38 (\pm IFN β) used as loading controls. **(E)** Total RNA was collected at the indicated times and measured for the indicated ISGs with *HINT1*, *PRDX6*, and *GUSB* used as loading controls.

(F–H) Acute CVB3 infection disrupts MAVS but not MDA5. **(F)** AC16-CAR cells were infected with CVB3 at MOI = 10 for the indicated times and immunoblotted for MDA5 (**F**, upper) or the full-length (FL) and mini-MAVS [mini; translated by leaky ribosomal scanning at Met¹⁴² (62)] with ectopic CAR-V5, p38, vinculin, and tubulin used as loading controls. **(G)** AC16-CAR cells were infected with CVB3 at MOI = 10 for 24 hours and immunoblotted for MDA5 or MAVS with vinculin, tubulin, and GAPDH used as loading controls. A 35 kDa MAVS cleavage product (MAVS₃₅) is visible 24 hours after infection. **(H)** Immunoblot densitometry of replicated experiments described in **(G)**.

(I) Mature virion formation predicted by complete kinetics with or without innate antiviral sensing (antiviral), viral antagonism of antiviral sensing (antagonism), or supplemental interferon (IFN) added at the indicated times. Predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

Data are shown as the mean \pm s.e.m. (**A**, **C**, **D**, **F**, **H**) or geometric mean \pm log-transformed standard error (**B**, **E**) of $n = 4$ biological replicates. For **(A–C)**, **(E)**, and **(F)**, time courses with significant alterations were assessed by one-way ANOVA (or, for MAVS_{FL} and MAVS_{mini}, two-way ANOVA) with replication, and a single asterisk indicates $p < 0.05$ for individual time points compared to $t = 0$ hour (gray band) after Tukey post-hoc correction. For **(D)** and **(H)**, a double asterisk indicates $p < 10^{-4}$ by Student's unpaired t test.

See also Figure 2.10.

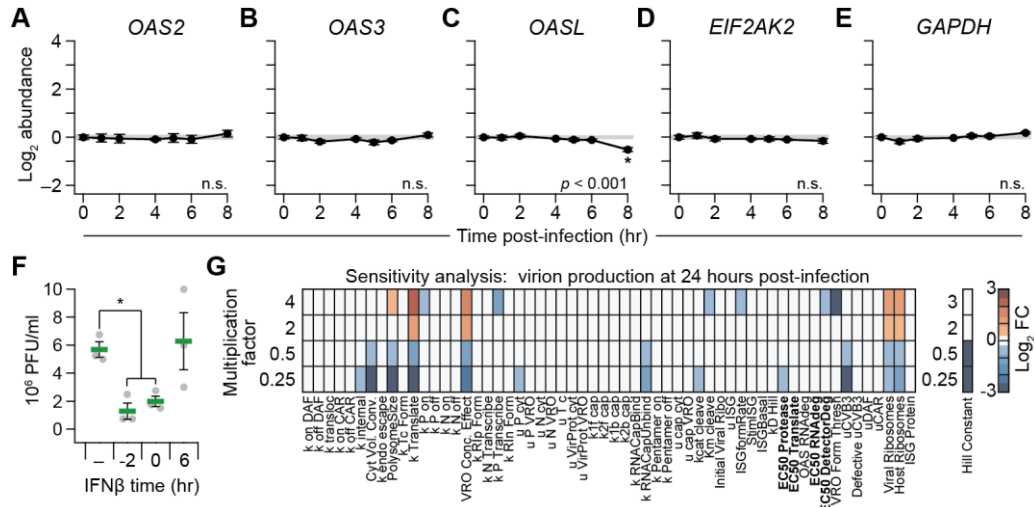


Figure 2.10. Controls for interferon-stimulated genes and time-dependent antagonism of mature virion formation by supplemental interferon.

(A–E) Total RNA was collected at the indicated times and measured for the indicated ISGs (*OAS2*, *OAS3*, *OASL*, *EIF2AK2*) or housekeeping gene (*GAPDH*) with *HINT1*, *PRDX6*, and *GUSB* used as loading controls. Data are shown as the geometric mean \pm log-transformed standard error of $n = 4$ biological replicates. Time courses with significant alterations were assessed by one-way ANOVA with replication, and a single asterisk indicates $p < 0.05$ for individual time points compared to $t = 0$ hour after Tukey post-hoc correction.

(F) Supplemental interferon blocks CVB3 progression during early infection. AC16-CAR cells were infected with CVB3 at MOI = 10 for 24 hours with 30 ng/ml IFN β added at the indicated times relative to the start of CVB3 infection at $t = 0$, and the conditioned medium was assessed by plaque assay. Data are shown as the mean PFU/ml \pm s.e.m. from $n = 3$ biological replicates.

(G) Terminal virion production is not sensitive to the exact choice of half-maximal effective concentrations (EC50s, bolded) specifying feedback weights. The listed model parameters were individually varied by the indicated multiplication factor and the concentration of virions reported at 24 hours as the log₂ fold change (FC) relative to the base parameter set. The Hill constant was varied multiplicatively below one and additively above one (where 2 = base Hill coefficient + 1).

2.3.6 A cleavage-resistant MAVS mutant shows enhanced antiviral activity upon delayed stimulation with paracrine interferon

We used the model to examine viral outcomes to delayed addition of interferon (Figure 2.10F). The scenario mimics a setting of paracrine antiviral signaling, where an infected cell is warned of a local infection by an interferon-secreting cell nearby. The earlier this paracrine warning signal is received, the more effective the stimulated interferon response can be at inhibiting viral propagation. We reasoned that the timed addition of interferon would uncover nonobvious combinatorial sensitivities in the network that could be tested experimentally. Using mature virions as the readout, we screened a five-tiered range of potencies for the three interferon feedbacks and viral-protease antagonism ($5^4 = 625$ simulations), and each simulation was run with or without interferon at five different times after infection [$5^4(1+5) = 3750$ simulations]. Many of the perturbations affected mature virions monotonically with effects that superposed when combined. However, the computational screen revealed a sensitivity to viral-protease antagonism that was dependent on delayed addition of interferons at intermediate times post-infection (Figures 2.11A and 2.11B). If interferon was supplemented early ($t \sim 4$ hr), viral propagation was dampened to a basal level regardless of the extent of antagonism. However, when viral-protease antagonism was reduced fivefold, the dampening persisted even when interferon was supplemented later ($t \sim 6$ hr). At this time, the base model cannot maintain peak ISG production because of excess viral protease but the resistant variant is still maximally fighting the infection (Figures 2.12A and 2.12B). The testable prediction was that perturbing sensor–transducer degradation by viral proteases would have analogous time-dependent effects when infected cells received supplemental interferons.

To test this prediction, we returned to the CVB3-induced cleavage of MAVS observed earlier (Figures 2.9G and 2.9H). MAVS can be cleaved by 3C^{pro} at a Gln-Ala cleavage site between positions 148 and 149 (5). However, such a cleavage product would be electrophoretically indistinguishable from MAVS_{mini}, an oligomerization-deficient proteoform that is translated by leaky ribosomal scanning at Met¹⁴² (62). The site's importance for CVB3 infectivity is also unclear, because Ala¹⁴⁹ in MAVS is undergoing rapid positive selection in primates (72). Chimpanzees (Val¹⁴⁹) and African green monkeys (Arg¹⁴⁹) both have substitutions that should prevent cleavage by 3C^{pro} (73), but coxsackievirus infections have been documented in both species (74, 75). We considered alternative sites that would be more consistent with the observed MAVS₃₅ product and identified a Gln-Gly cleavage site between positions 271 and 272. The site is included in a six amino-acid insertion distinguishing old-world monkeys/hominoids from new-world

monkeys, which are seronegative for coxsackievirus (76). The phylogenetics and observed cleavage pattern of MAVS built a stronger case for Gln²⁷¹ as an important site for sensor–transducer degradation during CVB3 infection.

Using a human MAVS template derived from a widely cited IMAGE clone (61, 77), we cloned a doxycycline-inducible, FLAG-tagged MAVS with or without Gln²⁷¹ substituted for alanine (Ala²⁷¹; Section 2.5, STAR Methods). Lentiviral transduction and selection of AC16-CAR cells yielded lines with comparably inducible ectopic MAVS alleles, which could be compared to one another (Figures 2.11C and 2.11D). When cells were induced and then infected with CVB3, we found that generation of MAVS₃₅ was significantly reduced in the Ala²⁷¹ line compared to the Gln²⁷¹ line for multiple viral stocks (Figures 2.11E and 2.11F). We attributed the residual MAVS₃₅ in the Ala²⁷¹ line to cleavage of endogenous MAVS at Gln²⁷¹. Thus, CVB3 antagonizes antiviral signal transduction by cleaving MAVS at Gln²⁷¹, and the Ala²⁷¹ line approximates a state of partial resistance as simulated by the complete kinetic model.

To evaluate the impact of MAVS cleavage on viral propagation, we collected medium from cells treated with CVB3 for 24 hours and titered infectious virions by plaque assay (Section 2.5, STAR Methods). Titers from the Ala²⁷¹ line were significantly lower relative to the Gln²⁷¹ line for multiple CVB3 stocks (Figures 2.11G and 2.11H). Biological error was usually close to the intrinsic counting noise of a plaque assay (Figures 2.12C and 2.12D), with the spread of biological replicates comparable to Poisson intervals about the mean (Figure 2.11H). We also kept in mind the increased uncertainty of the interferon-supplemented base model (Figure 2.11B, left) when comparing predictions to experiments. The Gln²⁷¹ and Ala²⁷¹ lines were next used to test for differences in the time-dependent effect of paracrine interferon signaling predicted by the model. We found that interferon- β treatment at 4 hours after infection significantly reduced the infectivity in the Gln²⁷¹ line, as paracrine stimulation offset the MAVS genotype (Figure 2.11I, left). At 5–6 hours, however, interferon- β lost efficacy in the Gln²⁷¹ line, returning to viral titers that were observed without paracrine stimulation. The Ala²⁷¹ line, by contrast, showed little, if any, time-dependent effect for interferon- β supplementation (Figure 2.11I, right). The Ala²⁷¹ mutant agreed even more closely with model predictions when using an inducible, cleavage-resistant MAVS allele containing a compound mutation at the other site reportedly cleaved by 3C^{pro} (Ala¹⁴⁸Ala²⁷¹; Figure 2.12E) (5). We conclude from these experiments that the overall importance of CVB3-proteinase susceptibility for MAVS is strongly modulated by paracrine interferon, consistent with the predictions of complete kinetics.

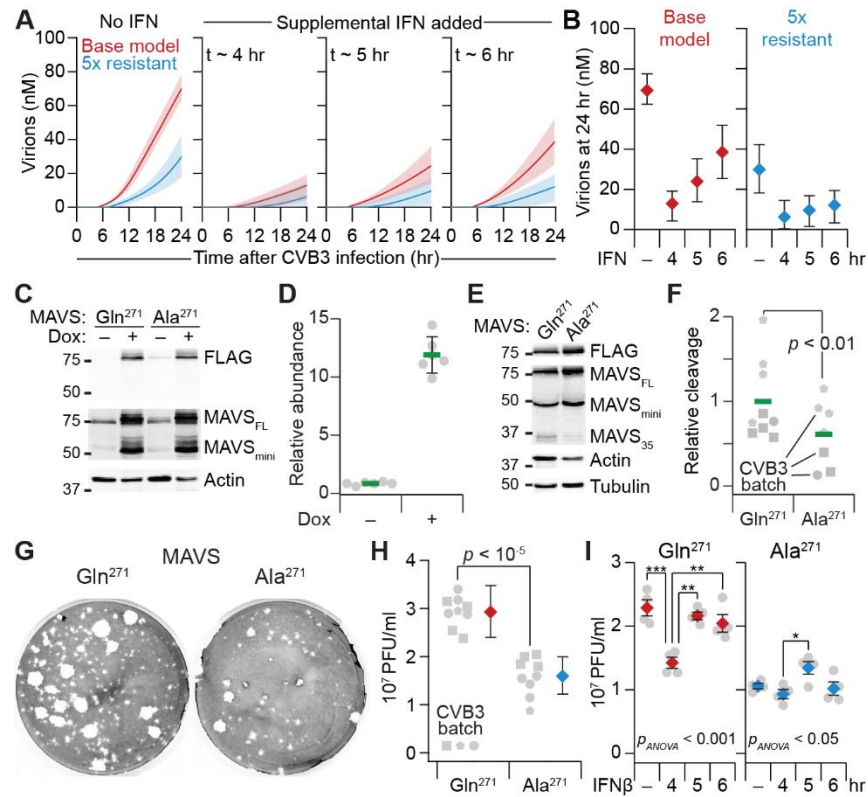


Figure 2.11. MAVS is a sensitive locus for CVB3 susceptibility in the host-cell network.

(A and B) CVB3 virion production is modulated nonlinearly by host-cell resistance and time-delayed supplementation of interferon. Virion output from complete kinetics with (blue) or without (red) fivefold altered resistance to viral proteinases. Negative feedback from the viral proteinases on the viral dsRNA sensor–transducer was decreased to mimic an increase in host-cell resistance. Supplemental interferon (IFN) was simulated at the indicated times. The 24-hour terminal endpoint of the simulations in (A) are summarized in (B). Predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

(C and D) Inducible ectopic expression of MAVS alleles. (C) AC16-CAR cells stably transduced with doxycycline (Dox)-regulated 3xFLAG-tagged Gln²⁷¹ MAVS or Ala²⁷¹ MAVS were treated with 1 μ g/ml Dox for 24 hours and immunoblotted for FLAG and MAVS with actin used as a loading control. The image gamma was adjusted for MAVS (gamma = 4) to show the endogenous full-length MAVS (MAVS_{FL}) and mini-MAVS (MAVS_{mini}) at the same exposure as the induced constructs. (D) Immunoblot densitometry of induced ectopic MAVS relative to endogenous MAVS_{FL} + MAVS_{mini} in AC16-CAR cells treated with or without Dox. Data are shown as mean \pm s.d. of $n = 6$ different MAVS alleles used in this work (see also Figure 2.13; Section 2.5, STAR Methods).

(E and F) CVB3-induced MAVS cleavage is reduced in cells expressing the Ala²⁷¹ mutant. (E) AC16-CAR cells stably transduced with inducible Gln²⁷¹ MAVS or Ala²⁷¹ MAVS were induced with Dox for 24 hours, then infected with CVB3 at MOI = 5 for 24 hours and immunoblotted for FLAG or MAVS with actin and tubulin used as loading controls. MAVS_{FL}, MAVS_{mini}, and the 35 kDa MAVS cleavage product (MAVS₃₅) are indicated. (F) Immunoblot densitometry of replicated experiments described in (E) using $n = 3$ different CVB3 batches and 1–4 biological replicates per batch.

(G and H) CVB3-induced virion release is reduced in cells expressing Ala²⁷¹MAVS. (G) Representative plaque assay for infectious virion release in the conditioned medium from AC16-CAR cells after induction of Gln²⁷¹ MAVS or Ala²⁷¹ MAVS and infection with CVB3 at MOI = 5 for 24 hours. (H) Quantification of plaque-forming units (PFU) from $n = 3$ different CVB3 batches and 1–4 biological replicates per batch. Data are summarized as PFU/ml \pm 95% Poisson confidence intervals based on the mean or the observation.

(I) Ala²⁷¹ MAVS sustains its antiviral potency upon delayed addition of beta-interferon (IFN β). After induction of Gln²⁷¹ MAVS or Ala²⁷¹ MAVS, AC16-CAR cells were infected with CVB3 at MOI = 10 for 24 hours, with 30 ng/ml IFN β added at the indicated times after the start of CVB3 infection. Data are shown as the mean \pm s.e.m. of $n = 4$ biological replicates, and differences across conditions was assessed by one-way ANOVA (p_{ANOVA}) with Tukey's posthoc test for individual differences: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The indicated reductions in viral titers between the two genotypes are qualitatively similar to a second Ala²⁷¹-harboring MAVS allele (Ala¹⁴⁸Ala²⁷¹) shown in Figure 2.12E.

For (F) and (H), the difference between MAVS genotypes was assessed by replicated two-way ANOVA with MAVS genotype and CVB3 batch as factors.

See also Figure 2.12.

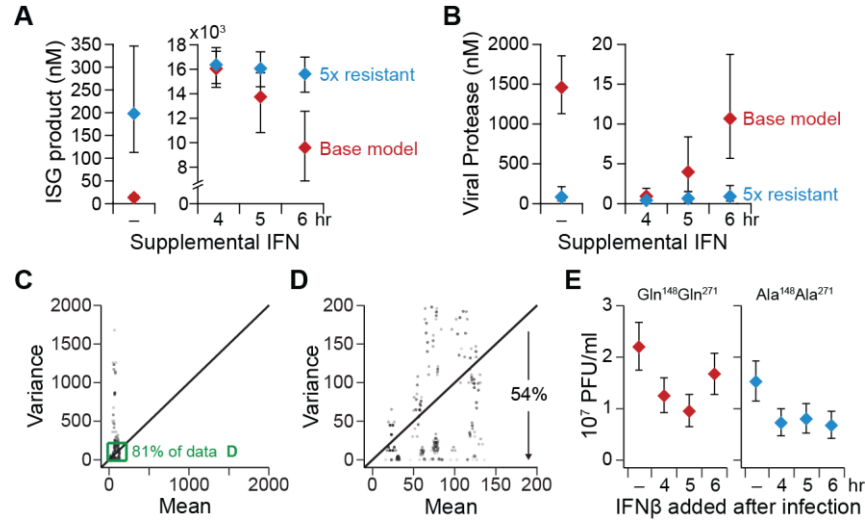


Figure 2.12. Poisson noise of the plaque assay and generality of the timed-interferon effect to Ala¹⁴⁸Ala²⁷¹-mutant MAVS.

(A and B) Predicted concentrations of ISG product and viral protease at 24 hours after infection with CVB3 and supplemental interferon (IFN) simulated at the indicated times. Predictions are shown as the median simulation \pm 90% nonparametric confidence interval from $n = 100$ simulations of single-cell infections at 10 PFU with a parameter coefficient of variation of 5%.

(C and D) Biological replicates of plaque assay are largely consistent with Poisson counting noise. After induction of LacZ or different MAVS alleles (Glu⁹³Gln¹⁴⁸Gln²⁷¹, Glu⁹³Gln¹⁴⁸Ala²⁷¹, or Gln⁹³Gln¹⁴⁸Gln²⁷¹; Section 2.5, STAR Methods), AC16-CAR cells were infected with CVB3 at MOI = 5 for 24 hours. Plaque assay data are from $n = 4$ biological replicates in one experiment (LacZ-induced cells), two experiments (Glu⁹³Gln¹⁴⁸Ala²⁷¹ and Gln⁹³Gln¹⁴⁸Gln²⁷¹ MAVS-induced cells), or three experiments (Glu⁹³Gln¹⁴⁸Gln²⁷¹ MAVS-induced cells). Each experiment was bootstrapped 1000 times, yielding 8000 mean–variance pairs. The inset in (C) capturing 81% of the bootstrap replicates (green) is expanded in (D).

(E) Ala¹⁴⁸Ala²⁷¹ MAVS qualitatively replicates Ala²⁷¹ MAVS antiviral potency upon delayed addition of beta-interferon (IFN β). After induction of Gln¹⁴⁸Gln²⁷¹ MAVS or Ala¹⁴⁸Ala²⁷¹ MAVS, AC16-CAR cells were infected with CVB3 at MOI = 5 for 24 hours, with 30 ng/ml IFN β added at the indicated times after the start of CVB3 infection. Data are from one plaque assay at $n = 4$ different IFN β conditions. Data are summarized as PFU/ml \pm 95% Poisson confidence intervals based on the observation.

2.3.7 Enteroviral proteinase cleavage of MAVS is redirected by a prevalent human polymorphism

The IMAGE clone used to construct the MAVS alleles is valid but contains three common polymorphisms that deviate from the reference protein sequence of humans and other mammals: Q93E (rs17857295; 25% aggregated allele frequency), Q198K (rs7262903; 16%), and S409F (rs7269320; 16%) (78). The Q198K polymorphism is considered functionally neutral (79), and none of these variants alter the mitochondrial localization of MAVS (80). However, the substitution at position 93 was relevant, because Gln⁹³ in the reference sequence creates the possibility of an additional site for cleavage by 3C^{pro} (Figure 2.13A) (73). Like Gln²⁷¹, cleavage at Gln⁹³ would separate the oligomerization domain (CARD) of MAVS and thereby prevent signal transmission from MDA5 to TRAFs on the mitochondrial surface (Figure 2.13A).

Considering the effect of the synthetic Ala²⁷¹ mutant (Figures 2.11C–I), we reasoned that position 93 might represent a naturally variable site affecting CVB3 infectivity in the human population. We replaced Glu⁹³ in the IMAGE clone sequence with Gln⁹³ and established AC16-CAR lines with equivalently inducible expression (Figure 2.11D). Upon CVB3 infection of the Gln⁹³ line, we detected a specific N-terminal MAVS fragment at ~18 kDa, which was consistent with cleavage at position 93 (Figures 2.13A and 2.13B). Given that the Gln⁹³ allele of MAVS harbors an additional 3C^{pro} site at Gln²⁷¹, we anticipated that this line would be even more permissive to CVB3 infection than the Glu⁹³ line. Unexpectedly, the opposite was true—the Gln⁹³ line yielded significantly fewer infectious virions than the Glu⁹³ line, despite twice as many 3C^{pro}-targetable sites (Figure 2.13C). Moreover, the Gln⁹³ allele of MAVS was cleaved much less extensively at position 271 in CVB3-infected cells, as indicated by the formation of MAVS₃₅ (Figure 2.13D). The results indicated a coupling between Gln⁹³ and Gln²⁷¹, where two cleavage options for 3C^{pro} rendered MAVS more resistant to viral antagonism than one.

There are now detailed models of MAVS activation triggered by dsRNA (81). However, before stipulating hierarchical or otherwise-special properties for different MAVS cleavages, we asked whether a separate, highly simplified model of MAVS regulation could be explanatory (Figure 2.13E). MAVS signaling occurs on the surface of mitochondria by polymeric self-assembly into filaments, which are nucleated by MDA5 binding to the dsRNA intermediates of CVB3 (11, 82). For the model, we assumed that MAVS activation occurs at a constant rate (k_{act}) and on the same time scale as the activation of 3C^{pro} (Section 2.5, STAR Methods). Structural data suggest that polymeric MAVS (polyMAVS) filaments are large (up to $n = 800$ molecules) (83, 84). We modeled polyMAVS formation as an

instantaneous assembly given that it occurs on a two-dimensional surface, and signaling from polyMAVS to ISGs was lumped as a first-order “transcription” process (k_{txn}). In cells, polyMAVS is degraded slowly by mitophagy (63), and we assumed a similar decay constant for ISGs (k_{deg} ; Figure 2.1C). These equations governed the behavior of the Glu⁹³Ala²⁷¹ allele that was assumed to be uncleavable by 3C^{pro}.

For cleavable alleles, we appended one (Glu⁹³Gln²⁷¹) or two (Gln⁹³Gln²⁷¹) first-order pathways in which 3C^{pro} slowly splits MAVS (k_{clv}) and removes it from the pool of monomers for polymerization (Clv1 or Clv2; Figure 2.13E). Either cleavage will remove the oligomerization domain from MAVS and render it unable to polymerize (Figure 2.13A). Recognizing that the remaining C-terminal truncations on the mitochondrial surface could further inhibit the polymerization of full-length MAVS (63), we allowed for negative feedback on the polymerization. In addition, we permitted unequal feedback between the two cleavage products for the Gln⁹³Gln²⁷¹ allele (Φ and $\gamma\Phi$), even though the simulations were largely insensitive to negative feedback overall (Figure 2.14A). When negative feedbacks are equal (Figure 2.13F) or absent (Section 2.6, Appendix, Note S1), the only difference between the simulations involving cleavage is that the polymerization-competent Gln⁹³Gln²⁷¹ is removed twice as quickly as Glu⁹³Gln²⁷¹.

In the filamentation model, all MAVS genotypes rapidly reach a pseudo-steady state as polyMAVS and ISGs begin to accumulate (Figure 2.13F). This trajectory is sustained for the Glu⁹³Ala²⁷¹ allele, because nothing abates the constant rate of MAVS activation (EA; Figure 2.13F). For the cleavable alleles, however, 3C^{pro} eventually reaches a concentration that depletes the reservoir of monomeric MAVS, causing polyMAVS to decline. When MAVS depletion reaches the point that polyMAVS formation is negligible compared to its mitophagic disposal, the kinetics of polyMAVS shift to a first-order decay with a long time constant (Section 2.6, Appendix, Note S1). For the Gln⁹³Gln²⁷¹ allele (QQ; Figure 2.13F), polyMAVS both decelerates and troughs twice as quickly, hence decaying slowly from a higher concentration of polyMAVS than that of the Glu⁹³Gln²⁷¹ allele (EQ; vertical arrows in Figure 2.13F). In this set of model parameters, the distinction is enough to change ISG kinetics from sustained to transient, but differences in ISG induction are preserved for a range of parameter sets (Figure 2.14A). The model illustrates how non-intuitive antiviral behavior arises when the kinetic competition between MAVS and 3C^{pro} is funneled through a surface-polymerization step that is highly nonlinear. If small differences in proteinase susceptibility are amplified by paracrine interferons (Figure 2.11), the minor-but-prevalent Glu⁹³ allele may thus contribute to the individual severity of human enteroviral infections.

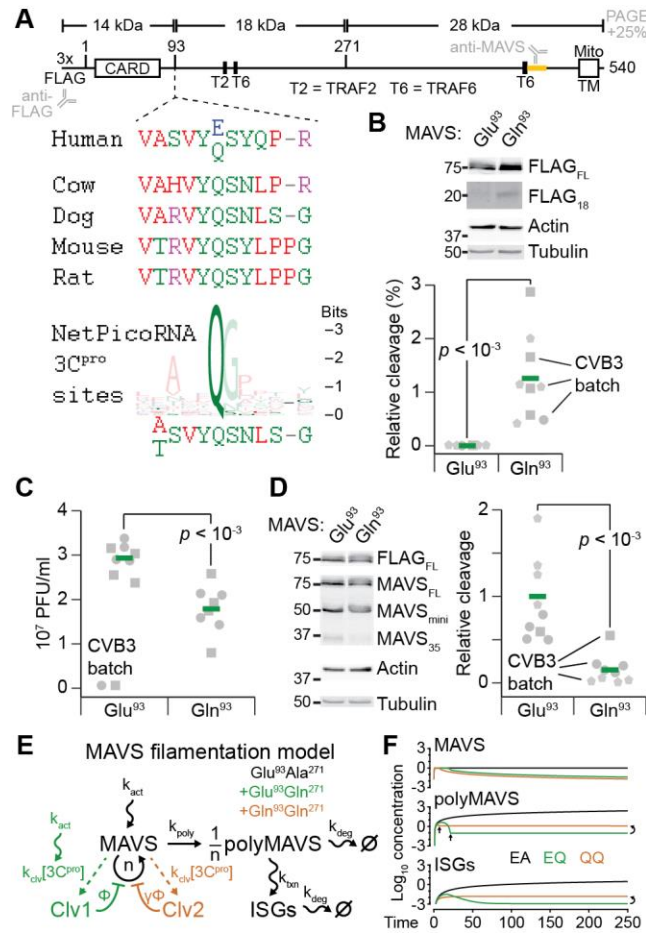


Figure 2.13. MAVS is a sensitive locus for CVB3 susceptibility in the human population.

(A) Sequence and domain architecture of MAVS. Above: predicted molecular weights of MAVS cleaved at the indicated positions—MAVS is acidic and separates on an SDS-polyacrylamide gel (PAGE) with an electrophoretic mobility ~25% larger than the predicted molecular weights listed. The oligomerization domain (CARD) and mitochondrial transmembrane (Mito TM) domains are indicated along with the recruitment sites for TRAF2 (T2) and TRAF6 (T6). The N-terminal epitope tag is indicated as well as the approximate peptide epitope of the anti-MAVS antibody (yellow). Below: flanking amino acids around position 93 of human MAVS, which has a Glu/Gln polymorphism. Gln⁹³ is widely conserved in mammals. The sequence logo for enteroviral 3C^{pro} was re-derived from the enteroviral cleavage sites analyzed by NetPicoRNA (Section 2.5, STAR Methods) (73).

(B) Ectopic expression of Gln⁹³ MAVS gives rise to an 18 kDa cleavage fragment. AC16-CAR cells stably transduced with inducible Glu⁹³ MAVS or Gln⁹³ MAVS were induced with doxycycline for 24 hours, then infected with CVB3 at MOI = 5 for 24 hours and immunoblotted for FLAG with actin and tubulin used as loading controls. The image gamma was changed for FLAG₁₈ (gamma = 4). (bottom) Immunoblot densitometry of replicated experiments using $n = 3$ different CVB3 batches and 1–4 biological replicates per batch. The difference between MAVS genotypes was assessed by replicated two-way ANOVA with MAVS genotype and CVB3 batch as factors.

(C) Ectopic Gln⁹³ MAVS significantly reduces virion release compared to ectopic Glu⁹³ MAVS. AC16-CAR cells stably transduced with inducible Glu⁹³ MAVS or Gln⁹³ MAVS were induced with Dox for 24 hours, then infected with CVB3 at MOI = 5 for 24 hours. Infectious virions in the conditioned medium were collected at the end of the 24-hour CVB3 infection. Quantification of plaque-forming units (PFU) is from $n = 2$ different CVB3 batches and four biological replicates per batch. The difference between MAVS genotypes was

assessed by replicated two-way ANOVA with MAVS genotype and CVB3 batch as factors.

(D) Ectopic expression of Glu⁹³ MAVS reduces the 35 kDa cleavage fragment. AC16-CAR cells stably transduced with inducible Glu⁹³ MAVS or Gln⁹³ MAVS were induced with Dox for 24 hours, then infected with CVB3 at MOI = 5 for 24 hours and immunoblotted for FLAG and MAVS with actin and tubulin used as loading controls (left). Immunoblot densitometry of replicated experiments using $n = 3$ different CVB3 batches and 1–4 biological replicates per batch (right). The difference between MAVS genotypes was assessed by replicated two-way ANOVA with MAVS genotype and CVB3 batch as factors.

(E) State-based model of MAVS self-assembly and cleavage by 3C^{pro}. The core model for the Glu⁹³Ala²⁷¹ mutant (black) is elaborated with one or two 3C^{pro}-catalyzed cleavage reactions for the Glu⁹³Gln²⁷¹ MAVS (green) or the Gln⁹³Gln²⁷¹ polymorphism (brown), respectively.

(F) Simulated trajectories for the three MAVS alleles: Glu⁹³Ala²⁷¹ (EA, black), Glu⁹³Gln²⁷¹ (EQ, green), and Gln⁹³Gln²⁷¹ (QQ, brown). Single-parameter sensitivity analysis for the time-integrated ISG profile is shown in Figure 2.14A, and the effect of MAVS genotypes in Figures 2.14B–D.

See also Figure 2.14 and Section 2.6, Appendix, Note S1.

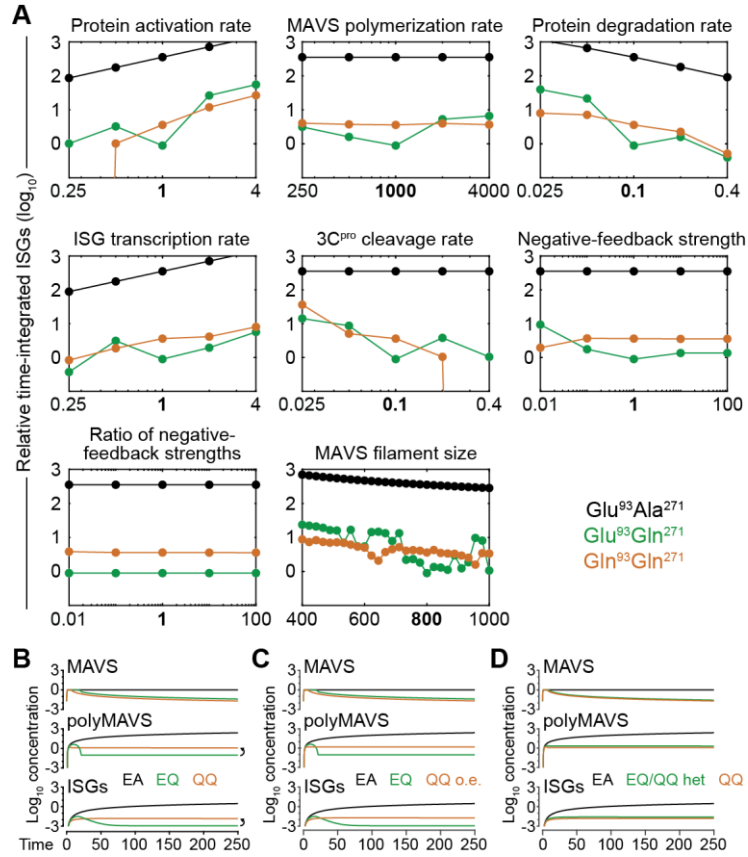


Figure 2.14. Sensitivity analysis of the MAVS regulation model.

(A) Individual model parameters were varied about their base value (bold) and simulations for the three genotypes repeated using integrated ISG abundance from 0–250 time units as the model output.

(B–D) Effect of mixed MAVS genotypes. The original simulations from Figure 2.13F are reprinted in (B) for comparison to (C) a Glu⁹³Gln²⁷¹/Gln⁹³Gln²⁷¹ heterozygote overexpressing Gln⁹³Gln²⁷¹ ~12-fold (QQ o.e., brown; Figure 2.11C) or (D) a Glu⁹³Gln²⁷¹/Gln⁹³Gln²⁷¹ heterozygote without overexpression (EQ/QQ het, green).

2.4 Discussion

This work examines the feasibility of complete kinetics for acute viral infections by leveraging the unsegmented genome of picornaviruses and 70+ years of enteroviral research in cells (85). We mathematically encoded the molecular pathways from viral binding and entry through the formation of mature virions. Interference by host-cell signaling was not ignored but superimposed as a set of feedbacks, which themselves were subject to increasing antagonism during viral progression. The resulting draft is consistent with measured parameters and other observations in the literature, as well as our own experiments. Collectively, the formalized mechanisms, computer simulations, and experimental results point to MAVS as a critical determinant of the enteroviral response in human cells.

Generalization to other enteroviruses and cellular contexts

Organizing the complete kinetics of CVB3 into modules is advantageous (Figure 2.1A), because it will streamline adaptation to other enteroviruses. Within the genus, the most substantive differences lie in binding and entry. Poliovirus, for instance, uses a single cell-surface receptor with two binding affinities and delivers genome to the cytoplasm before it is fully endocytosed (2, 19, 86). Exchange of a poliovirus-specific delivery module may be sufficient to explain the accelerated kinetics of infection relative to CVB3.

For any enterovirus species, results from complete kinetics suggest that the particle-to-PFU ratio is a critical determinant for predicting host-cell tropism and susceptibility. Viruses with large ratios of defective particles will depend on excess cell-surface receptors to infect, which may not be true in all humans. Receptors for coxsackievirus decline considerably with age (87), presumably to different extents among individuals. For CAR, the interquartile range of transcript expression is 1.7–5.6 TPM (4000–10,000 copies per cell) in the left ventricle (22), which straddles the inferred threshold for productive infection with the CVB3 stocks used here (~7000 copies per cell; Figures 2.3E and 2.4F). The delivery module can accommodate such host-cell variation in a scalable and principled way.

Different VROs for different kinetic regimes of viral infection

All positive-strand RNA viruses reconfigure host-cell membranes to promote replication (36). However, the membrane rearrangements differ among viruses, and it is unclear whether function of the resulting VROs is the same. Complete kinetic modeling of CVB3 indicates that the main advantage conferred by enteroviral VROs is to accelerate replication biochemistry on membrane surfaces. The hydrophobic 3AB protein binds 3D^{pol}

and supports its intrinsic propensity to form two-dimensional lattices (54). In turn, VROs emerge right at the onset of exponential enterovirus replication (40).

A computational model of flavivirus reached a different conclusion about the role of VROs (38). Parameter sensitivity of VROs in this model emphasized the importance of compartmentalizing RNA-dependent RNA polymerase and shielding the positive-strand template from degradation. The apparent discrepancy can be reconciled when considering the different rates of infection and replication between the two RNA-virus genera. Flaviviral polymerase associates >60-fold more slowly and synthesizes at one-third the rate of enteroviruses (88) (Table S1), with infections requiring multiple days to establish. The much slower time scales should increase the importance of sequestering flavivirus RNA with polymerase and away from host-cell ribonucleases. Indeed, flaviviral VROs assemble as membranous webs or invaginations that shield replication components from the host (36), supporting kinetics as a pressure for viral evolution.

Expanding and revising complete kinetic models

We emphasize that the complete kinetics of CVB3 reported here is a draft that should be subject to future refinements. Many reaction parameters were drawn from literature on poliovirus (Table S1), which is more similar to certain A-type coxsackievirus strains than to CVB3 (89). Also, the physical interactions between viral RNA and the VRO surface were not specified with the same detail as elsewhere in the kinetic model. If the longer-lived 3CD^{pro} precursor confers specific RNA-binding properties to the membrane (90), the stoichiometry of this intermediate may need to be considered explicitly. The threshold for VROs was the one instance where a model parameter was “tuned” to data. Although uncertainty may persist around the switch to VRO-like surface behavior, future alternatives will be easy to vet because VRO initiation is such a critical transition in the simulations.

The feedbacks overlaid on the kinetic model could certainly be elaborated more deeply, but there are also advantages to lumping at this scale. For example, MAVS is cleaved at a different site (Gln⁴²⁸) by the mitochondrially localized 3ABC precursor of hepatitis A picornavirus (91). The net result, however, is the same as cleavage at Gln⁹³ or Gln²⁷¹—separation of the MAVS oligomerization domain from the mitochondrial surface where it can polymerize rapidly. A different challenge relates to incorporating the human MAVS alleles with different cleavage susceptibilities. Although the simplified MAVS regulatory model is intriguing, it is premature to incorporate as part of complete kinetics. The modeled step is but one in a pathway, using parameters that were nominal approximations of fast and slow processes. Even within this step, the true length of

polymerized MAVS filaments in cells is actively debated (92). Frankly, cell signaling is not really geared for complete kinetics like fast-acting RNA viruses are. For cell signaling, more appropriate are the abstractions and flexible system boundaries of models that are incomplete but useful (93).

Implications for enteroviral disease

The Q93E polymorphism in MAVS occurs with a mutant allele frequency of ~50% among individuals of East Asian ancestry (78), indicating a large proportion of homozygous individuals in these populations. Geographically, many Asian countries face widespread cyclical outbreaks of enterovirus infection, and the understanding of risk factors is incomplete (94). The MAVS regulatory model suggests that a pure Glu⁹³ context behaves very differently from a mixture of Glu⁹³ and Gln⁹³ (Figures 2.14B–D). Based on the MAVS RNA-seq reads, we gleaned that AC16 cells are heterozygous for the Q93E allele. Our results argue for more-formal studies of CVB3 susceptibility in isogenic lines that investigate different genotypes of MAVS expressed at endogenous levels.

Pharmacologically, 3C^{pro} has long been recognized as a therapeutic target for enteroviral infections, with inhibitors reaching as far as Phase II (95–97). The rationale for such compounds is to block maturation of the enteroviral polypeptide, but it can be challenging to target an intramolecular cleavage with sufficient potency in cells. Intermolecular 3C^{pro} activity toward MAVS provides a means for reappraisal, especially considering its variable cleavability in humans. According to the encoded viral feedback, a 10–20-fold shift in the EC₅₀ for sensor–transducer degradation is sufficient to block a low-titer CVB3 infection (MOI = 1) entirely.

Complete kinetics is a tangible organizing principle for viruses with a limited gene repertoire and an acute mode of infection. The concept may seem a distant goal for large viruses with multiple unknown gene products. It is nonetheless a goal that can specify where operational paradigms are unsatisfactory. Overall, viruses are much more modular than the host cells they infect. Systems bioengineers should exploit this property to define the integrated parts lists that could one day be mixed and matched to propose complete kinetic models for viruses that have recently recombined (18, 98).

2.5 STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kevin A. Janes (kjanes@virginia.edu).

Materials Availability

Plasmids related to this work are deposited with Addgene (#158628–158646).

Data and Code Availability

- RNA-seq source data have been deposited at the NCBI Gene Expression Omnibus and are publicly available under the accession number: GSE155312.
- MATLAB original code for the complete kinetic model of CVB3 is publicly available at <https://github.com/JanesLab/CompleteKinetics-CVB3>. MATLAB original code for the MAVS filamentation model is publicly available at <https://github.com/JanesLab/MAVSfilamentation>.
- The scripts used to generate the modeling figure subpanels reported in this paper are available at <https://github.com/JanesLab/LopacinskiAB-CellSyst2021>. Scripts were not used to generate the experimental figure subpanels reported in this paper.
- Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

AC16 (27) and AC16-CAR cells (28) were cultured in DMEM/F12 (Gibco) + 12.5% fetal bovine serum (Hyclone) + 1% penicillin/streptomycin (Gibco) and kept at 37°C, 5% CO₂.

METHOD DETAILS

Plasmids

The lentiviral destination vector pLX302 (Addgene #25896) was recombined with pDONR221 EGFP (Addgene #25899) or pDONR223 MX1, OAS1, OAS2, and OASL from the Human ORFeome v5.1 (99) to yield pLX302 EGFP-V5 puro (Addgene #158644), pLX302 MX1-V5 puro (Addgene #158640), pLX302 OAS1-V5 puro (Addgene #158641), pLX302 OAS2-V5 puro (Addgene #158642), and pLX302 OASL-V5 puro (Addgene #158643). pDONR223 MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ was obtained from the human ORFeome v5.1 (99),

originally derived from IMAGE clone #5751684. MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ amplicon was prepared with SpeI and MfeI restriction sites and cloned into pEN_TTmiRc2 3xFLAG (Addgene #83274) that had been digested with SpeI and MfeI (Addgene #158628). pDONR223 MAVS Glu⁹³Ala¹⁴⁸Gln²⁷¹ was obtained by site-directed mutagenesis of pDONR223 MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ (99). MAVS Glu⁹³Ala¹⁴⁸Gln²⁷¹ amplicon was prepared with SpeI and MfeI restriction sites and cloned into pEN_TTmiRc2 3xFLAG (Addgene #83274) that had been digested with SpeI and MfeI (Addgene #158630). pEN_TT-3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Ala²⁷¹ (Addgene #158631) and pEN_TT-3xFLAG-MAVS Gln⁹³Gln¹⁴⁸Gln²⁷¹ (Addgene #158629) were obtained by site-directed mutagenesis of the pEN_TT-3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ plasmid (Addgene #158628). pEN_TT-3xFLAG-MAVS Glu⁹³Ala¹⁴⁸Ala²⁷¹ (Addgene #158633) was obtained by site-directed mutagenesis of the pEN_TT-3xFLAG-MAVS Glu⁹³Ala¹⁴⁸Gln²⁷¹ plasmid (Addgene #158630). pEN_TT-3xFLAG-MAVS Gln⁹³Gln¹⁴⁸Ala²⁷¹ (Addgene #158632) was obtained by site-directed mutagenesis of the pEN_TT-3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Ala²⁷¹ plasmid (Addgene #158631). Site-directed mutagenesis of pEN_TT-3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Ala²⁷¹ consistently caused one of two off-target mutations. Two clones with different off-target mutations were digested with MfeI and NaeI. The fragments were gel purified, and the fragments with the non-off-target mutations were ligated together to yield pEN_TT-3xFLAG-MAVS Gln⁹³Gln¹⁴⁸Ala²⁷¹ (Addgene #158632). All site-directed mutagenesis was performed with the QuikChange II XL kit (Agilent).

pEN_TT donor vectors were recombined into pSLIK hygro by LR recombination to obtain pSLIK 3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ hygro (Addgene #158634), pSLIK 3xFLAG-MAVS Glu⁹³Ala¹⁴⁸Gln²⁷¹ hygro (Addgene #158636), pSLIK 3xFLAG-MAVS Glu⁹³Gln¹⁴⁸Ala²⁷¹ hygro (Addgene #158637), pSLIK 3xFLAG-MAVS Gln⁹³Gln¹⁴⁸Gln²⁷¹ hygro (Addgene #158635), pSLIK 3xFLAG-MAVS Glu⁹³Ala¹⁴⁸Ala²⁷¹ hygro (Addgene #158639), and pSLIK 3xFLAG-MAVS Gln⁹³Gln¹⁴⁸Ala²⁷¹ hygro (Addgene #158638).

DNA for CVB3 was obtained by partial digestion of CVB3-M1 (100) (kindly provided by K. Klingel) with EcoRI, gel purification of the full-length CVB3 insert, and ligation into pcDNA3 digested with EcoRI to obtain pcDNA3 CVB3 (Addgene #158645). The shRNA targeting SV40 T antigen (GCATAGAGTGTCTGCTATTAA) was custom designed through the Genetic Perturbation Platform web portal (<https://portals.broadinstitute.org/gpp/public/>). The oligos were modified to contain a PstI-containing loop, self-annealed, and ligated into pLKO.1 neo (Addgene #13425) to yield pLKO.1 shSV40 neo (Addgene #158646).

Viral Transduction

pSLIK MAVS alleles, pSLIK LacZ, pSLIK Luc, and pLKO.1 shSV40 were packaged as lentivirus in HEK293T cells by calcium phosphate precipitation with an EGFP-expressing plasmid as a positive control. AC16-CAR cells were transduced with lentivirus + 8 µg/ml polybrene in 6-well dishes and allowed to grow for 48 hours. Transduced pSLIK cells were then transferred to 10-cm dishes and selected with 10 µg/ml blasticidin (to maintain CAR-V5 expression) and 100 µg/ml hygromycin until control plates had cleared. Transduced shSV40 cells were used unselected.

AC16 Quiescence

For studies involving quiescent cells (Figures 2.3B, 2.3C, and 2.4B–D only), AC16 or AC16-CAR cells (70,000 per well) were plated on a 12-well dish precoated with 0.02% (w/v) gelatin for 1–2 hours at 37°C. After 24 hours, cells were washed with PBS and the culture medium was changed to differentiation medium: DMEM/F12 (Gibco) + 2% horse serum (Thermo Fisher) + 1x Insulin–Transferrin–Selenium supplement (Thermo Fisher) + 1% penicillin–streptomycin (Gibco). After 24 hours, cells were washed with PBS and transduced with shSV40 lentivirus (125 µl per well) + 8 µg/ml polybrene in a total volume of 500 µl. After 18–21 hours, cells were washed with PBS and refed with differentiation medium. After another 24 hours, cells were washed with PBS and refed with differentiation medium again, and cells were lysed 24 hours after the final refeed.

CVB3 Infection

AC16-CAR cells were plated at 50,000 cells/cm² for 24 hours. AC16-CAR cells expressing inducible MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹, Glu⁹³Gln¹⁴⁸Ala²⁷¹, Glu⁹³Ala¹⁴⁸Ala²⁷¹, or Gln⁹³Gln¹⁴⁸Gln²⁷¹ or LacZ were plated at 25,000 cells/cm² for 24 hours then treated with 1 µg/ml doxycycline for 24 hours. Before CVB3 infection, 75% of the cell culture medium was removed, and cells were infected with CVB3 at the indicated MOI for one hour. During the infection, the plates were incubated at 37°C and rocked every 10–15 minutes to ensure even coverage of the virus. After one hour, the media was aspirated, and cells were refed with fresh AC16 growth medium lacking selection antibiotics. At the end of the infection period, the conditioned media containing released virions was collected, centrifuged at 2,500 rcf to spin out dead cells and debris, and stored at –80°C for viral titering.

Cell Morphometry

Cell dimensions were analyzed by staining terminally CVB3-infected cells as before (101). Briefly, AC16-CAR cells were infected with CVB3 at 10 MOI for 24 hours, stained with

the LIVE/DEAD Fixable Violet Dead Cell Stain Kit (Thermo Fisher), quenched with 10 mM glycine in PBS for five minutes, solvent fixed–permeabilized with 100% ice-cold methanol, and counterstained with DRAQ5 (Cell Signaling Technology). Imaging at late times after CVB3 infection allowed VROs to fuse and grow above the diffraction limit for fluorescence imaging (37). One hundred Violet-positive cells were imaged with an Olympus BX51 fluorescence microscope with a 40x 1.3 numerical aperture UPlanFL oil immersion objective and an Orca R2 charge-coupled device camera (Hamamatsu) with 2x2 binning. Images were thresholded in ImageJ to identify Violet-positive membrane borders, Violet-negative VRO borders, and DRAQ5-positive nuclear borders. When calculating cellular dimensions, we assumed a height equal to the diameter of a sphere corresponding to the average aggregate VRO volume per cell.

Cell Lysis

For immunoblotting, cells were washed with ice-cold PBS and lysed in radioimmunoprecipitation buffer plus protease–phosphatase inhibitors: 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% (v/v) Triton X-100, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA, 10 µg/ml aprotinin, 10 µg/ml leupeptin, 1 µg/ml pepstatin, 1 µg/ml microcystin-LR, 200 µM Na₃VO₄, and 1 mM PMSF. Protein concentrations of clarified extracts were determined with the BCA Protein Assay Kit (Pierce).

For RNA analysis, culture medium was aspirated and cells were immediately lysed in Buffer RLT Plus. Total RNA was purified with the RNEasy Mini Plus kit (Qiagen) as recommended. RNA concentrations were determined by absorption spectrophotometry on a Nanodrop.

Immunoblotting

Quantitative immunoblotting was performed on 10, 12, or 15% polyacrylamide gels with tank transfer to polyvinylidene difluoride membrane and multiplex near-infrared fluorescence detection as described previously (102). Primary antibodies were used at the following dilutions: VP1 (Mediagnost, 1:2000 dilution), eIF4G (Cell Signaling Technology, 1:2000 dilution), V5 epitope tag (Invitrogen, 1:5000 dilution or Bethyl, 1:5000 dilution), HSP90 (Santa Cruz Biotechnology, 1:2000 dilution), tubulin (Abcam, 1:20,000 dilution or Cell Signaling Technology, 1:2000 dilution), p38 (Santa Cruz Biotechnology, 1:5000 dilution), vinculin (Millipore, 1:10,000 dilution), IκBα (Cell Signaling Technology, 1:2000 dilution), phospho-TBK1 (Cell Signaling Technology, 1:1000 dilution), total TBK1 (Cell Signaling Technology #3504), phospho-IRF3 (Cell Signaling Technology, 1:1000 dilution), total IRF3 (Cell Signaling Technology, 1:1000 dilution), MDA5 (Abcam, 1:1000 dilution),

MAVS (Cell Signaling Technology, 1:1000 dilution), actin (Ambion, 1:5000 dilution), FLAG (Sigma-Aldrich, 1:5000 dilution). Serial dilutions of V5-containing Multiple Tag (GenScript) were used for absolute quantification of CAR-V5.

In Vitro Transcription of CVB3 Positive and Negative Strands

The pcDNA3 CVB3 plasmid was linearized at the 3' end with NotI (positive-strand template) or at the 5' end with SnaBI (negative-strand template), gel purified, and used for in vitro transcription with T7 (positive strand) or SP6 (negative strand) RNA polymerase from the MAXIscript SP7/T7 kit (Ambion). In vitro transcription reactions were incubated at 37°C for two hours before the template was digested with TURBO DNase I at 37°C for 15 minutes and transcripts purified by LiCl precipitation according to the manufacturer's recommendations (Thermo Fisher). RNA concentrations were determined by absorption spectrophotometry on a Nanodrop.

dsRNA Transfection

Poly(I:C) (High Molecular Weight) and 5'ppp-dsRNA control were prepared according to the manufacturer's separate recommendations (Invivogen). dsCVB3 was prepared by mixing equal volumes of CVB3 positive and negative strand at 100 ng/μl in 0.9% (w/v) NaCl, denaturing for at 68°C for 10 minutes, and cooling to 25°C at 0.1°C per second on a thermocycler.

For dsRNA transfection, AC16-CAR cells were plated at 50,000 cells/cm² on a 12-well plate overnight and transfected with up to 1 μg dsRNA complexed with 3 μl Lipofectamine 3000 + no P3000 reagent (Invitrogen) in 200 μl total volume of DMEM/F12 (Gibco). For dose responses, lipocomplexes were serially diluted threefold in DMEM/F12 over ~two decades before addition to AC16-CAR cells. Four hours after transfection, cells were lysed in 350 μl Buffer RLT Plus and purified with the RNEasy Plus Mini Kit (Qiagen). Purified RNA concentrations were determined by absorption spectrophotometry on a Nanodrop.

ISG Transfection and Protein-Synthesis Inhibition

To estimate ISG cellular half-lives, 293T/17 cells were plated at 100,000 cells/cm² on a 24-well plate and transfected with 1 μl Lipofectamine 3000 + 1 μl P3000 reagent (Invitrogen) and 250 ng pLX302 EGFP-V5 puro + 250 ng pLX302 ISG-V5 puro (ISG = MX1, OAS1, OAS2, or OASL) in 100 μl total volume of DMEM (Gibco). Twenty-four hours after transfection, cells were treated with 50 μM cycloheximide (Sigma-Aldrich) for the indicated times and lysed in 50 μl RIPA + PPIs for immunoblot analysis of ~30 μg extract for V5

epitope tag (Invitrogen, 1:5000 dilution [MX1-V5] or Bethyl, 1:5000 dilution [OAS1-V5, OAS2-V5, and OASL-V5]) and HSP90 (Santa Cruz Biotechnology, 1:2000 dilution)–tubulin (Cell Signaling Technology, 1:2000 dilution)–p38 (Santa Cruz Biotechnology, 1:5000 dilution) as loading controls. Cotransfection of pLX302 EGFP-V5 yielded a short-lived FP-V5 truncation that served as a positive control for cycloheximide efficacy (FP_{control}).

RNA-seq and Analysis

500 ng total RNA was prepared using the TruSeq Stranded mRNA Library Preparation Kit (Illumina). Samples were sequenced on a NextSeq 500 instrument with NextSeq 500/550 Mid Output v2 kit (150 cycles; Illumina). Pooled samples were run in triplicate to obtain a minimum sequencing depth of 20 million reads per sample. Adaptors were trimmed using fastq-mcf in the EAutils package (version 1.1.2) with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, and no nucleotide skew causing cycle removal). Datasets were aligned to the human (GRCh38) genome with additions (CVB3 genome and SV40 genome) using HISAT2 (version 1.2.0) with the following options: --dta (downstream transcriptome assembly for subsequent assembly step) and --rna-strandedness RF (for paired-end reads generated by the TruSeq strand-specific library). Output SAM files were converted to BAM files using samtools (version 1.4.1). Alignments were assembled into transcripts using StringTie (version 2.0.6) with the following options: -e (to restrict assembly to known transcripts in the provided annotation) and -B (to save additional files for Ballgown). Differential gene expression analysis was carried out using edgeR (version 3.28.1) on raw read counts that passed the abundance-filtering step. Abundance filtering was performed by the cpm function in edgeR to retain transcripts that were expressed at greater than 100 counts per million in at least one cell line. Trimmed mean of M values normalization using the calcNormFactors function before differential expression analysis using exactTest in edgeR. The 1952 transcripts that were commonly differentially expressed [5% false discovery rate (FDR)] between AC16 parental and AC16-CAR cells are shown in Figure 2.3B.

HeLa Cell Transcriptomic and Proteomic Data

Raw HeLa cell transcriptomic data (GSE111485) was downloaded from the GEO sequence read archive. The data were aligned and TPM calculated exactly as for the AC16 and AC16-CAR cells. “HeLa12” and “HeLa14” were profiled in triplicate, so each of their TPM values were averaged. Proteomic data was obtained from <https://helaprot.shinyapps.io/crosslab/> by searching for CD55 for DAF and CXADR for CAR. The 14 HeLa samples were paired by numeric index with the 14 HeLa TPM samples

obtained from GEO. Obtained proteomic data was in log₁₀ format, so absolute protein abundances were calculated by $10^{(\log_{10_protein})}$.

Determination of CVB3 Particle-to-PFU Ratio

Titered production stocks of CVB3 prepared in permissive HeLa cells were ultracentrifuged at 100,000 rcf for one hour at 4°C and lysed in Buffer RLT Plus for total RNA purification with the RNEasy Mini Plus kit (Qiagen) as recommended. RNA concentrations were determined by absorption spectrophotometry on a Nanodrop. Conditioned medium from uninfected cells was used as a negative control to confirm specificity to viral RNA. Similar results were obtained when CVB3 was ultracentrifuged at 120,000 rcf for 18 hours at 8°C through a 30% sucrose cushion and into a glycerol button (103).

Quantitative PCR (qPCR)

Purified total RNA was digested with TURBO DNase (Thermo Fisher) as recommended, and 500 ng of DNase-treated RNA was reverse transcribed with 2.5 µM oligo(dT)₂₄ primer as described previously (104). *PRDX6*, *HINT1*, and *GUSB* were used as loading controls, and *GAPDH* was used as a fourth housekeeping gene to confirm accuracy of the loading normalization.

Tagged Strand-Specific qPCR

Purified total RNA was digested with TURBO DNase (Thermo Fisher) as recommended, and 250 ng of DNase-treated RNA was reverse transcribed with 1 pmol biotinylated strand-specific primer (positive strand: 5'biotin-GGGTGTTCTTTGGATCCTTG; negative strand: 5'biotin-TGCAACTCCCATCACCTGTA), 500 µM dNTPs, 5 mM DTT, 20 U RNasin Plus RNase inhibitor (Promega), and 100 U Superscript III reverse transcriptase (Invitrogen) in 1x first-strand buffer as recommended in a total volume of 10 µl. After reverse transcription for 60 minutes at 55°C, samples were heat inactivated for 15 minutes at 70°C and RNA was digested with 2.5 U RNase H (NEB) for 20 minutes at 37°C, followed by heat inactivation for 20 minutes at 65°C. Purified standards—either RNA from CVB3 virions purified by ultracentrifugation (positive strand) or in vitro transcribed RNA from SnaBI-digested pcDNA3 CVB3 (negative strand)—were added in the range of 10⁸–10³ copies per 10 µl reverse transcription along with 250 ng of DNase-treated RNA from cells that had not been infected with CVB3.

Biotinylated cDNA was purified with 5 µl streptavidin magnetic beads (Thermo Fisher) that had been washed twice with 1x first-strand buffer and resuspended to the

original storage volume. Streptavidin beads were incubated with the first-strand reaction for one hour with tapping every 15 minutes. The 15 μ l samples were diluted with 40 μ l PCR-grade water, magnetized for one minute, and then washed three times with 50 μ l PCR-grade water. After the final wash, the beads were resuspended in 50 μ l and then 1/10th of the nonbead volume (4.5 μ l) was measured by qPCR with 2.5 pmol each of CVB3-specific primers (Table S2) as described previously (104).

Plaque Assay

Conditioned medium from CVB3-infected AC16-CAR cells was collected and diluted 5×10^4 or 5×10^5 in serum-free DMEM depending on the titer. A confluent 6-well plate of permissive HeLa cells was washed twice with serum-free DMEM, then 200 μ l of diluted conditioned medium or serum-free DMEM was added to the cells. The cells were incubated at 37°C for one hour with rocking every 10–15 minutes to ensure even coverage of the virus. After one hour, the medium was aspirated and cells were washed twice with serum-free DMEM. Sterile 1.5% (w/v) agar in water was mixed with 2x DMEM in a 1:1 ratio and added to the wells to form an agar plug. The cells were incubated at room temperature 10–15 minutes until solidification of the agar plug. Cells were incubated at 37°C for 48–65 hours depending on the clearance rate of the plaques. Once plaques reached a countable size, 2 ml of 2% (w/v) formaldehyde was added to each well for 15 minutes to fix the cells. After fixation, the agar plugs were carefully pulled out using a plastic spatula. Fixed cells were stained with 0.5% (w/v) crystal violet in 25% (v/v) methanol in Milli-Q water for 10 minutes. Crystal violet staining was stopped by submerging the plates three times into a 1% bleach solution, and then wells were gently rinsed with Milli-Q water three times from a squirt bottle. The plates were air dried and scanned on a Licor Odyssey scanner in the 700 channel.

3C^{pro} Sequence Logo

We recreated the sequence logo for enteroviral 3C^{pro} substrates by assembling the 202 3C^{pro} cleavage sites from 27 SwissProt accession numbers in Table 6 of the original publication describing NetPicoRNA (73). Flanking sequences were amended with the latest revisions in UniProt, and the vectorized sequence logo was generated with WebLogo (105).

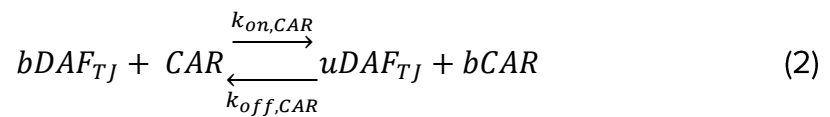
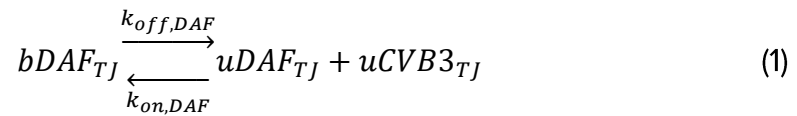
CVB3 Complete Kinetic Model

The complete kinetics of CVB3 infection were described by 54 coupled differential equations that were organized into modules for delivery, replication, and encapsidation.

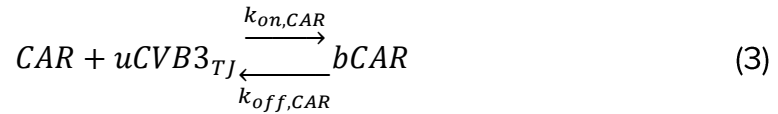
After construction, the modules were interconnected by viral inputs–outputs and host-cell feedback. The system of differential equations was solved numerically with the ode15s function in MATLAB. The two core functions are 1) CVB3ODEEval.m, which defines the initial conditions and rate parameters, takes the user-defined inputs (CVB3 dose, simulation parameters, output options), and performs the bookkeeping on viral species (RNA strands, replicative intermediates, and protein classes); and 2) CVB3ODEfunc.m, which defines the rate equations, the switch to VROs, and the antiviral sensing-transduction mechanism. Additional functions provide capabilities for systematic plotting and sensitivity analysis (Data and Code Availability). MATLAB version 2019b or later is required to handle the export of tabular data, and the Statistics Toolbox is required to display sensitivity analyses. The following Method Details explain and justify the key assumptions within each module, feedback, and bookmarked species.

Delivery Module

At the start of a single-cell infection, a user-defined number of PFUs was placed in the restricted cell-surface volume defined by the estimated cell surface area multiplied by the 30-nm diameter of a CVB3 capsid (Table S1). For a cell-population infection, cell-to-cell variation in response to a viral inoculum follows a Poisson distribution (20). Therefore, a Poisson-distributed pseudorandom number (representing a single-cell PFU) was drawn at each simulation with λ equal to the user-defined MOI. Cellular DAF and CAR were also concentrated in the restricted cell-surface volume for CVB3 binding and trafficking. Clustering of DAF was embedded in the rate of DAF trafficking to tight junctions, which used a kinetic parameter drawn from observations that relocalization is complete within 25–30 minutes (25). Tight junctions were not encoded as a volume-restricted compartment but rather as their own set of species on the plasma membrane. DAF-bound CVB3 in tight junctions ($bDAF_{TJ}$) was allowed to unbind-rebind in the tight junction or transfer CVB3 directly to CAR:



In addition, CAR could bind CVB3 that had dissociated in tight junctions ($uCVB3_{TJ}$):



The forward rates for Equations 2 and 3 were assumed to be identical to place equal weight on the two possible paths to CAR-bound CVB3. CAR internalizes with CVB3 (26) and we assumed that it was not recycled; however, this assumption was inconsequential to any of the kinetics (Figure 2.4A).

Entry of CVB3 occurs more slowly than poliovirus (19, 26). Therefore, delivery of CVB3 to the cytoplasm was encoded as two first-order processes in series, the first using a standard endocytosis rate and the second capturing pore formation and release of viral RNA (Table S1). Defective viral particles were assumed to bind to DAF–CAR and deliver to the cytoplasm identically to infectious PFUs.

Replication Module

Mammalian cells contain 10^6 – 10^7 ribosomes (106). Considering competition with newly transcribed host-cell RNAs and with RNA from defective particles, we assumed that 10^5 ribosomes per cell were theoretically accessible to the infectious virus. This number is also roughly equivalent to the number of eIF4G molecules [$\sim 10^{5.5}$ copies per HeLa cell (30), with TPM values comparable in AC16 cells], allowing us to equate eIF4G–ribosome as a single species in the model. We further simplified by considering viral translation in terms of polysomes containing N ribosomes. Each polysome translated viral polyprotein N times more rapidly but formed and released a translational complex with viral RNA N times more slowly than an individual ribosome. Translational initiation was emphasized as a rate-limiting step by selecting an operational early-polysome size of $N = 2.5$ (Figure 2.6A). Defective viral RNAs in the cytoplasm were assumed not to interact with the ribosome pool that was theoretically accessible, but this assumption was not critical to infection outcomes (Figures 2.6E and 2.6F).

Among these polysomes, 80–90% would be actively translating host-cell RNAs and therefore be inaccessible at the time of infection (106). The conversion of polysomes from inaccessible to accessible was encoded as an enzyme-catalyzed transition rate dependent on the concentration of viral protease, the concentration of inaccessible ribosomes (i.e., uncleaved eIF4G), and Michaelis-Menten parameters for rhinoviral 2A^{pro} (107) (Table S1). The reported K_M for enterovirus 71 3C^{pro} is equivalently far above endogenous substrates and shows comparable catalytic efficiency (108), enabling 2A^{pro}–3C^{pro} to be lumped as a single viral protease.

The dissociation of a translation complex is non-intuitive and thus explained in detail here. The rate of protein production used poliovirus translation rates per amino acid multiplied by the length of the CVB3 polyprotein (109). Each translation event yields one 3D^{pol} bound to VRO, one lumped protease, $\frac{1}{5}$ of a capsid pentamer at the VRO (which couples to $\frac{1}{5}$ of a 2C^{ATPase}), and $\frac{4}{5}$ of unbound 2C^{ATPase} at the VRO to balance polyprotein stoichiometry. At a rate N^{-1} times the translation rate, the translation complex dissociates to release a positive-strand RNA at the VRO (33, 34) and repopulate the available polysome pool by N ribosomes. Placing a translation-competent positive strand at the VRO mimics cis replication (33, 34), in that defective viral RNAs do not reach this step in the complete kinetic model. More-direct alternatives, such as coupling 3D^{pol} with released positive-strand RNA as a replicative intermediate at the VRO, gave rise to premature increases in negative strand that were not observed by experiment (Figure 2.5C).

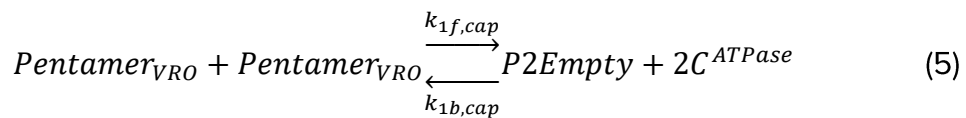
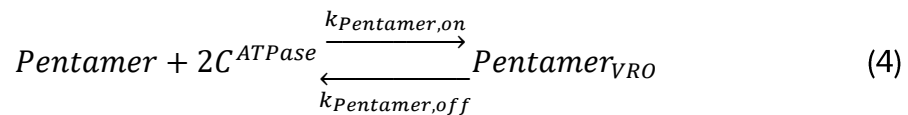
The early steps of VRO formation remain elusive and were specified heuristically in the complete kinetic model. We assumed a switch-like transition from solution-phase behavior in the cytoplasm to the restricted surface volume of the VRO defined by the estimated VRO surface area multiplied by the 7.1-nm height of a 3D^{pol} molecule (Table S1). Based on these calculations, the concentrating effect (C_E) of shuttling a molecule from the cytoplasm to the VRO surface volume was 3216-fold. Degradation rates at the VRO surface were reciprocally scaled down by C_E to yield the same turnover as in the cytoplasm. The switch-like transition to VROs at 25 molecules of 3D^{pol} created a transient stoichiometric imbalance at the start of the transition (Figure 2.8B). The shift occurred when the rate processes of VRO-resident species were instantly scaled up and their whole-cell concentrations were reciprocally scaled down. The imbalances were short-lived, and alternate encodings of the VRO transition, such as a ramped transition, did not satisfactorily capture the measured protein–RNA dynamics.

At the VRO, viral RNA interacted with 3D^{pol} at the slow rate of association measured for poliovirus 3D^{pol} (110). The rate of RNA-dependent RNA polymerase elongation was likewise drawn from poliovirus (21). We did not postulate any differences in association or elongation between positive- and negative-strand templates (111). However, we assumed that negative strand was not released from a replicative intermediate after replication of the positive strand was complete and had been released. This assumption reflects the cooperative RNA binding that results from 3D^{pol} oligomeric arrays on the VRO surface (54). Positive strands, by contrast, must be released for encapsidation.

Encapsulation Module

The encapsidation module leveraged the work of Zlotnick and colleagues for modeling viral capsid assembly (56, 112, 113). During assembly, contact affinities that are weak individually (~ 1 mM) build upon one another through multivalency among symmetric subunits with multiple binding sites (113). Using established formalisms (112), we calculated effective pseudocritical concentrations (apparent dissociation constants) for all realistic combinations of pentamer and contact sites in a 12-pentamer virus assembly. We took the median effective pseudocritical concentration (~ 1 μ M, calculated with Capsid_assembly_script.m) as a representative affinity for any step during the assembly process from 1–12 pentamers. By the same analysis, a slight increase in the individual contact affinity (0.1 mM; Figure 2.7E) increased the median affinity to 50 nM (simulated by decreasing the off rate; Figure 2.7E). The simplification avoids the need to assume or inventory specific geometric paths to viral assembly while recognizing the typical binding energies involved.

Recruitment and retention of pentamers to the VRO occurred through a direct interaction of 1:1 stoichiometry with $2C^{ATPase}$ (50). Kinetic parameters for this interaction are not known, but the complete kinetic model was not very sensitive to a biologically plausible range of values (Figure 2.7F). Crucial to the stoichiometric balance was to keep track of each molecule of $2C^{ATPase}$ “consumed” by pentamer binding to the VRO and then “regenerated” when pentamers self-assembled. For example:



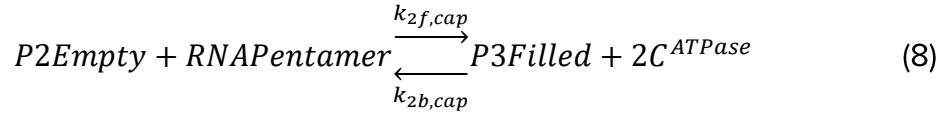
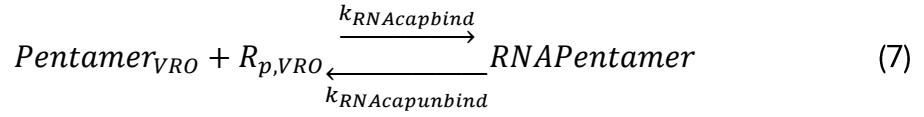
Pentamer assemblies were assumed to grow and shrink linearly, with individual pentamer or RNA–pentamer complexes recruited to or removed from an intermediate assembly.

The mechanisms retaining viral RNA in the VRO are heterogeneous, requiring a more-generic exchange rate between cytoplasmic and VRO compartments. The on and off rates were equal ($k_{on} = k_{off} = 1$ hr⁻¹), but at the VRO, C_E (= 3216) provided a driving force for RNA to leave that was delayed by the retention time implied by the rate of exchange:

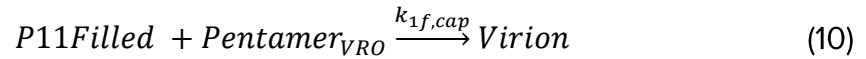
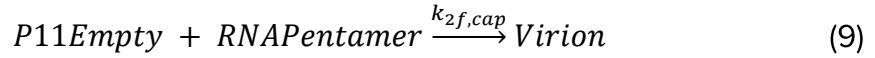


At these exchange rates, a typical RNA-protein association rate of 25 nM⁻¹hr⁻¹ (55) yielded an effective membrane affinity of (3216 × 1 hr⁻¹)/25 nM⁻¹hr⁻¹ ~ 125 nM.

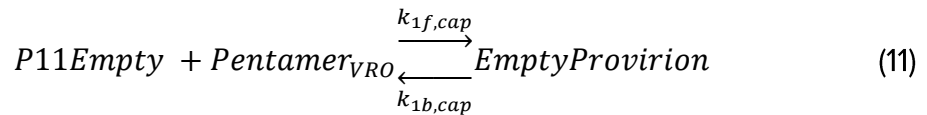
RNA and pentamer interacted at the VRO with the same kinetic parameters as the median pentamer–pentamer assemblies described above. We reduced combinatorics by assuming that RNA filling–unfilling of an intermediate assembly only occurred through gain or loss of an RNA–pentamer complex, which was defined to exist only at the VRO surface. For example:



(Note the regenerated 2C^{ATPase} in Equation 8, consistent with the stoichiometric balance described in Equation 5.) The assembly of 12 pentamers with positive-strand RNA was considered the irreversible endpoint of encapsidation, which could be reached two ways:



Empty provirions were assumed to be reversible:



like the other steps in the encapsidation module. Rate parameters involving $Pentamer_{VRO}$ ($k_{1f,cap}$ and $k_{1b,cap}$) were assumed to be identical to those involving $RNAPentamer$ ($k_{2f,cap}$ and $k_{2b,cap}$).

Antiviral and Antagonistic Feedbacks

The feedbacks overlaid on the CVB3 life cycle reflected known antiviral and antagonistic mechanisms, but they were not mechanistically encoded. Rather than striving for exact parameter values, we focused on lumped parameters and relationships that appropriately reflected dependencies and input–output characteristics. Future revisions could include more-elaborate detail when warranted by specific applications.

For the dsRNA sensor–transducer feedback leading to an interferon response, we defined a virus detection sigmoid with parameters (K_D , n_H) gleaned from the *IFNB* induction of cells lipofected with dsCVB3 (Figure 2.1B). To capture viral antagonism, the sigmoid maximum was scaled down hyperbolically depending on the concentration of viral protease and a half-effective concentration ($EC50_{detectordeg}$) defined by qualitative system properties (Figure 2.9I):

$$VirDetection = \left(\frac{[dsRNA]^{n_H}}{[dsRNA]^{n_H} + K_D^{n_H}} \right) \left(\frac{EC50_{detectordeg}}{[Protease] + EC50_{detectordeg}} \right) \quad (12)$$

where $[dsRNA]$ is the total concentration of replicative intermediates (Bookkeeping). This sigmoid was used to scale a maximum induced response of ISGs, which was defined by the maximum transcription rate for RNA polymerase II through a short ISG, multiplied by the current number of virus-inaccessible ribosomes (Table S1). For simulations involving supplemental interferon, *VirDetection* was superseded by *StimISG* = 1 (maximum stimulated interferon response) at the time of supplementation and maintained for the rest of the simulation.

The ISG response impinges on three facets of the viral life cycle. For i) the inhibition of viral translation, we hyperbolically down-scaled the formation rate of translation complexes:

$$TranslateForm' = TranslateForm \left(\frac{EC50_{Translate}}{[ISG] + EC50_{Translate}} \right) \quad (13)$$

Although PKR (*EIF2AK2*) is highly abundant (~480 TPM) and does not show interferon inducibility in AC16-CAR cells (Figure 2.2A), PKR is directly activated by dsRNA. Thus, the feedback architecture was assumed to be functionally similar to other ISGs. For ii) the deactivation of viral proteinases, we used a hyperbolic down-scaling function (similar to that of Equation 13) on the rate of viral protease production, assuming that protease deactivation was irreversible. For iii) the oligoadenylate synthetase–RNAse L acceleration

of RNA turnover, we posited a maximum fold increase in the rate of turnover ($OAS_{RNAdeg} = 5$) and scaled the basal rate of RNA degradation (k_{RNAdeg}) hyperbolically to this maximum based on a half-effective concentration ($EC50_{RNAdeg}$) and the concentration of ISGs:

$$k_{RNAdeg}' = k_{RNAdeg} \left[1 + (OAS_{RNAdeg} - 1) \left(\frac{[ISG]}{[ISG] + EC50_{RNAdeg}} \right) \right] \quad (14)$$

All feedback parameters were sampled lognormally about their best estimate with a user-defined log coefficient of variation as with the parameters and initial conditions of the complete kinetic model.

Bookkeeping

Total dsRNA concentrations combined the replicative intermediates using positive- or negative-strand templates. For ratiometric calculations where dsRNAs were counted with different efficiencies (Figure 2.5E), only the template strand was included in the calculation. Total viral protein included capsid from infectious PFUs and defective particles from internalization and all steps downstream (Figure 2.3A). Viral particles bound to DAF or CAR on the cell surface were assumed to be removed by the washing steps associated with cell lysis and protein analysis. Endosomal release of R_p implied instant degradation of all associated capsid proteins. For RNA analysis, cells were not washed before lysis according to the manufacturer's recommendation; therefore, these calculations included all viral particles bound to CAR on the cell surface (DAF affinity is so low that it was assumed to be removed by aspiration).

Sensitivity Analysis

Single-parameter sensitivity analysis was performed with SensitivityAnalysisfunc.m (Data and Code Availability) to alter individual model parameters over a 2²-fold range in either direction from the base value and record virion production at 24 hours.

MAVS Filamentation Model

The filamentation model for MAVS was described by 4–6 differential equations depending on the MAVS genotype:

$$\frac{d[MAVS]}{dt} = k_{act} - k_{poly}[MAVS]^n - \alpha k_{clv}[3C^{Pro}][MAVS] \quad (15)$$

$$\frac{d[polyMAVS]}{dt} = k_{poly}[MAVS]^n - \frac{k_{deg}}{n} \frac{\varphi}{\varphi + \frac{clv1}{\gamma} + clv2} [polyMAVS] \quad (16)$$

$$\frac{d[3C^{Pro}]}{dt} = k_{act} \quad (17)$$

$$\frac{d[ISG]}{dt} = \frac{k_{txn}}{n} [polyMAVS] - k_{deg}[ISG] \quad (18)$$

$$\frac{d[Clv1]}{dt} = (\alpha - \beta)k_{clv}[3C^{Pro}][MAVS] \quad (19)$$

$$\frac{d[Clv2]}{dt} = \beta k_{clv}[3C^{Pro}][MAVS] \quad (20)$$

where $\alpha = 0$, $\beta = 0$ for Glu⁹³Ala²⁷¹, $\alpha = 1$, $\beta = 0$ for Glu⁹³Gln²⁷¹, and $\alpha = 2$, $\beta = 1$ for Gln⁹³Gln²⁷¹. For the Gln⁹³Gln²⁷¹/Glu⁹³Gln²⁷¹ heterozygous model (Figure 2.14D), $\alpha = 1.5$, $\beta = 0.5$, and for the Gln⁹³Gln²⁷¹ overexpression model in a Gln⁹³Gln²⁷¹/Glu⁹³Gln²⁷¹ heterozygous background (Figure 2.14C), α was set to $1 + 12.5/13$ and β to $12.5/13$ to reflect the ~12-fold ectopic expression measured in Figure 2.11C.

Initial conditions were all set to zero, and rate parameters were nominally assumed to be fast ($k_{poly} = 1000$), average (k_{act} , $k_{txn} = 1$), or slow (k_{clv} , $k_{deg} = 0.1$). Φ and γ were set to one as equal default feedback strengths, and $n = 800$ based on calculations from the literature (83, 84). Filament assembly was assumed to occur by a two-step process used for other biological filaments (114). Dynamic trajectories were solved numerically for 250 nominal time units with the ode15s function in MATLAB. Single-parameter sensitivities were evaluated by changing parameters individually 2²- or 10²-fold in either direction about the nominal parameter and using time-integrated ISG abundance as the model readout.

QUANTIFICATION AND STATISTICAL ANALYSIS

dsRNA-Induced Interferon Response

Dose-independent induction of *IFNB*, *MX1*, and *OAS1* as a group (or *PKR* and *GAPDH* as a control group) was assessed by log-transformed three-way ANOVA of ISG transcripts with the following factors: dsRNA dose, transcript, and condition (control vs. poly(I:C) or control vs. dsCVB3). ANOVA *p* values testing a condition effect were Šidák corrected for multiple-hypothesis testing.

ISG Protein Half-Life Determination

Immunoblot bands were quantified by image densitometry as described previously (102). V5 band intensities were normalized to the averaged proportional loadings of

HSP90, tubulin, and p38 for each sample on the immunoblot. Loading-normalized biological duplicates were averaged, and the ISG and FP_{control} time courses were normalized again by their respective maxima. Half-lives ($t_{1/2}$) were estimated by nonlinear least-squares curve fitting in Igor Pro (WaveMetrics) to the following:

$$\text{Max-normalized protein} = (1 - b)e^{-kt} + b \quad (21)$$

Where b represents a minimum baseline (estimated from the FP_{control} of each experiment and kept constant for the associated ISG) and $t_{1/2} = \ln(2)/k$.

Plaque Assay Calculations

Plaques were counted manually in ImageJ after contrast enhancement. The plaque forming units per milliliter (PFU/ml) was calculated from the following formula:

$$PFU/ml = \frac{(\# \text{ plaques})(\text{dilution})}{\text{volume of virus added in ml}} \quad (22)$$

The dilution was either 5×10^4 or 5×10^5 , and the volume of virus added to the wells was 0.2 ml.

Relative Cleavage of MAVS Genotypes

MAVS cleavage was assessed via densitometry in ImageJ as

$$\% \text{ cleavage} = \frac{MAVS_{clv}}{MAVS_{clv} + MAVS_{mini} + MAVS_{FL}} \quad (23)$$

and normalized to MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ such that MAVS Glu⁹³Gln¹⁴⁸Gln²⁷¹ cleavage = 1. Differences in cleavage were assessed by two-way ANOVA with CVB3 batch and genotype as factors.

Protein Abundance Estimation From TPM Data

Paired transcriptomic-proteomic data in HeLa cells (30) were related to one another using the following hyperbolic-to-linear equation:

$$\text{Protein copies} = A \left(\frac{TPM}{B + TPM} + TPM \right) \quad (24)$$

where A and B are fitted parameters. Equation 24 specifies a linear relationship between

RNA and protein (29) but allows for some nonlinearity at low RNA copy numbers. Nonlinear least-squares parameterization was performed with the `fminsearchbnd` function in MATLAB. Asymptotic error analysis was used to calculate the 99% confidence interval of the fit. DAF and CAR protein copies were estimated in AC16 and AC16-CAR cells using the best-fit from the HeLa data for DAF or CAR, respectively.

2.6 Appendix

Note S1. Behavior of MAVS alleles in the MAVS filamentation model.

The single-parameter sensitivity analysis indicated that time-integrated ISG profiles are robust to feedback inhibition of MAVS cleavage products on the uncleaved MAVS polymerization rate (Figure 2.14A). Feedback loops are thus omitted from this analysis of dynamical behavior, defined by a system of 4–6 differential equations depending on the MAVS allele:

Glu ⁹³ Ala ²⁷¹	Glu ⁹³ Gln ²⁷¹	Gln ⁹³ Gln ²⁷¹
$\frac{d[M]}{dt} = k_a - k_p[M]^n$ (1a)	$\frac{d[M]}{dt} = k_a - k_p[M]^n - k_c[P][M]$ (1b)	$\frac{d[M]}{dt} = k_a - k_p[M]^n - 2k_c[P][M]$ (1c)
$\frac{d[pM]}{dt} = k_p[M]^n - \frac{k_d}{n}[pM]$ (2)	$\frac{d[pM]}{dt} = k_p[M]^n - \frac{k_d}{n}[pM]$ (2)	$\frac{d[pM]}{dt} = k_p[M]^n - \frac{k_d}{n}[pM]$ (2)
$\frac{d[P]}{dt} = k_a$ (3)	$\frac{d[P]}{dt} = k_a$ (3)	$\frac{d[P]}{dt} = k_a$ (3)
$\frac{d[I]}{dt} = \frac{k_t}{n}[pM] - k_d[I]$ (4)	$\frac{d[I]}{dt} = \frac{k_t}{n}[pM] - k_d[I]$ (4)	$\frac{d[I]}{dt} = \frac{k_t}{n}[pM] - k_d[I]$ (4)
	$\frac{d[C_1]}{dt} = k_c[P][M]$ (5)	$\frac{d[C_1]}{dt} = k_c[P][M]$ (5)
		$\frac{d[C_2]}{dt} = k_c[P][M]$ (6)

where M is uncleaved MAVS, pM is polymerized MAVS filaments, P is 3C^{Pro}, C_1 is MAVS cleavage product 1, C_2 is MAVS cleavage product 2, I is ISGs, k_a is a general activation rate, k_p is the polymerization rate, k_d is a general degradation rate, k_t is the transcription rate, and n is number of MAVS monomers in a polyMAVS filament. For the same rate parameters, differences in the magnitude and duration of I depend entirely on the dynamics of pM , which depends on the dynamics of M . Further, the linear accumulation of P allows its analytical solution ($k_a t$) to be substituted directly into the governing equation for M . Ignoring the bookkeeping equations of C_1 and C_2 reduces the system to two stable nonlinear, nonhomogeneous differential equations:

$$\frac{d[M]}{dt} = k_a - k_p[M]^n - \alpha k_c k_a t [M] \quad (7)$$

$$\frac{d[pM]}{dt} = k_p[M]^n - \frac{k_d}{n}[pM] \quad (8)$$

where $\alpha = 0$ (Glu⁹³Ala²⁷¹), 1 (Glu⁹³Gln²⁷¹), or 2 (Gln⁹³Gln²⁷¹).

For $\alpha = 0$, the long-term dynamics of the system can be solved analytically, with M rapidly reaching its pseudo steady-state value (M_{ss}):

$$M_{ss} = \sqrt[n]{\frac{k_a}{k_p}}. \quad (9)$$

Substituting M_{ss} into the governing equation for pM yields:

$$\frac{d[pM]}{dt} = k_a - \frac{k_d}{n} [pM] \quad (10)$$

and the solution:

$$[pM](t) = \frac{k_a n}{k_d} \left(1 - e^{-\frac{k_d}{n} t} \right) \quad (11)$$

with pM approaching $\frac{k_a n}{k_d}$ slowly with time, because $n \gg k_d$.

When $\alpha \neq 0$, an analytical solution does not exist. However, it is possible to identify critical transitions and limiting behavior of the dynamical system. For example, because t is always linearly increasing, one can define the time at which this initial M_{ss} is no longer possible and M must decrease. Changing Equation (7) to an inequality:

$$k_a - k_p [M]^n - \alpha k_c k_a t < 0 \quad (12)$$

and solving for t yields a transition time for M (t_M):

$$t_M > \frac{k_a - k_p [M]^n}{\alpha k_c k_a} \quad (13)$$

Thus, if the state of M is equal, the model predicts that $\text{Gln}^{93}\text{Gln}^{271}$ ($\alpha = 2$) will transition twice as fast as $\text{Glu}^{93}\text{Gln}^{271}$ ($\alpha = 1$; Figure SN1).

Immediately after t_M , all terms in Equations (7) and (8) contribute to the dynamical-system behavior. However, the high multiplicity of MAVS polymerization implies that M will eventually decrease to a point that $k_p [M]^n$ rapidly vanishes relative to the other terms in $\frac{d[M]}{dt}$.

$$\frac{d[M]}{dt} \approx k_a - \alpha k_c k_a t [M] = k_a (1 - \alpha k_c t [M]) \quad (14)$$

At this point, both M and t are large enough that we can further approximate Equation (14) as:

$$\frac{d[M]}{dt} \approx -\alpha k_c k_a t [M] \quad (15)$$

which is solvable analytically:

$$M(t) \approx c_1 e^{-\frac{\alpha k_c k_a}{2} t^2} \quad (16)$$

where c_1 is an unknown coefficient. Note from Equation (16) that the exponential decline occurs half as fast for Glu⁹³Gln²⁷¹ ($\alpha = 1$) compared to Gln⁹³Gln²⁷¹ ($\alpha = 2$). However, the approximate time point at which the $k_p[M]^n$ term would vanish—shortly after t_M —is twice as long for Glu⁹³Gln²⁷¹ according to Equation (13). Therefore, around the time when the system begins to deviate from steady state, the net decrease in M (and thus pM) occurs twice as fast for Glu⁹³Gln²⁷¹ as for Gln⁹³Gln²⁷¹ (Figure SN1).

Finally, at longer times, the rapid decline in M according to Equation (16) has progressed so much that it overwhelms the offsetting linear increase in t in Equation (14), and the simplification in Equation (15) is no longer valid. Here, the dynamics of M are governed by the linear, nonhomogeneous Equation (14). Numerical simulations suggest it is around this transition when $k_p[M]^n$ vanishes relative to $\frac{k_d}{n}[pM]$ and the governing equation for pM [Equation (8)] can be approximated as:

$$\frac{d[pM]}{dt} \approx -\frac{k_d}{n}[pM] \quad (17)$$

and solved to yield:

$$pM(t) \approx c_2 e^{-\frac{k_d}{n}t} \quad (18)$$

where c_2 is an unknown coefficient. In this regime, pM does not reach a formal steady state but is operationally stable at c_2 because $n \gg k_d$ (Figure SN1). For Glu⁹³Gln²⁷¹, the delayed transition from steady state [Equation (13)] causes a faster rate of decline in M [Equation (16)], leading to a deeper plunge in pM before the transition to an operationally stable concentration [Equation (18)].

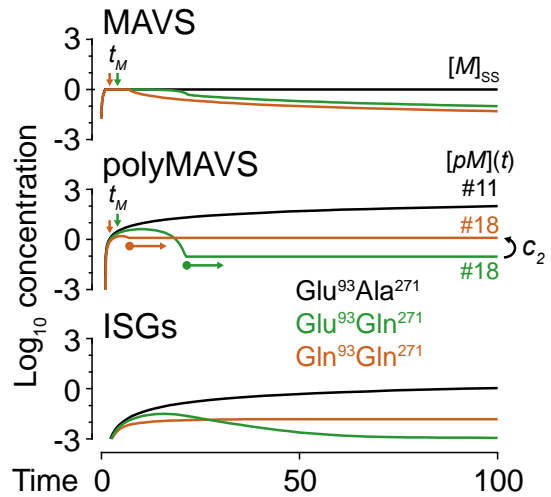


Figure SN1. Annotated dynamics of the MAVS filamentation model. See text for description of variables and equations.

2.7 References

1. J. W. Moore, R. G. Pearson, *Kinetics and Mechanism* (John Wiley & Sons, 1981).
2. V. Racaniello, “Picornaviridae: the viruses and their replication” in *Fields Virology*, 6th ed, D. M. Knipe, P. M. Howley, Eds. (Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013).
3. J. Yin, J. Redovich, Kinetic Modeling of Virus Growth in Cells. *Microbiol Mol Biol Rev* **82** (2018).
4. M. Esfandiarei, B. M. McManus, Molecular biology and pathogenesis of viral myocarditis. *Annu Rev Pathol* **3**, 127–155 (2008).
5. A. Mukherjee, *et al.*, The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog* **7**, e1001311 (2011).
6. J. M. Bergelson, *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320–1323 (1997).
7. D. R. Shafren, D. T. Williams, R. D. Barry, A decay-accelerating factor-binding strain of coxsackievirus B3 requires the coxsackievirus-adenovirus receptor protein to mediate lytic infection of rhabdomyosarcoma cells. *J Virol* **71**, 9844–9848 (1997).
8. J. Baggen, H. J. Thibaut, J. R. P. M. Strating, F. J. M. van Kuppeveld, The life cycle of non-polio enteroviruses and how to target it. *Nat Rev Microbiol* **16**, 368–381 (2018).
9. A. Friedman, N. Perrimon, Genetic screening for signal transduction in the era of network biology. *Cell* **128**, 225–231 (2007).
10. J. P. Wang, *et al.*, MDA5 and MAVS mediate type I interferon responses to coxsackie B virus. *J Virol* **84**, 254–260 (2010).
11. Q. Feng, *et al.*, MDA5 detects the double-stranded RNA replicative form in picornavirus-infected cells. *Cell Rep* **2**, 1187–1196 (2012).
12. J. W. Schoggins, *et al.*, A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
13. R. Deonarain, D. Cerullo, K. Fuse, P. P. Liu, E. N. Fish, Protective role for interferon-beta in coxsackievirus B3 infection. *Circulation* **110**, 3540–3543 (2004).
14. M.-L. Liu, *et al.*, Type I interferons protect mice against enterovirus 71 infection. *J Gen Virol* **86**, 3263–3269 (2005).
15. Q. Feng, *et al.*, Enterovirus 2Apro targets MDA5 and MAVS in infected cells. *J Virol* **88**, 3369–3378 (2014).
16. A. Haghighat, *et al.*, The eIF4G-eIF4E complex is the target for direct cleavage by the rhinovirus 2A proteinase. *J Virol* **70**, 8444–8450 (1996).

17. D. Etchison, S. C. Milburn, I. Edery, N. Sonenberg, J. W. Hershey, Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 220,000-dalton polypeptide associated with eucaryotic initiation factor 3 and a cap binding protein complex. *J Biol Chem* **257**, 14806–14810 (1982).
18. K. A. Janes, *et al.*, An engineering design approach to systems biology. *Integr Biol (Camb)* **9**, 574–583 (2017).
19. B. Brandenburg, *et al.*, Imaging poliovirus entry in live cells. *PLoS Biol* **5**, e183 (2007).
20. E. L. Ellis, M. Delbrück, THE GROWTH OF BACTERIOPHAGE. *J Gen Physiol* **22**, 365–384 (1939).
21. C. Castro, *et al.*, Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci U S A* **104**, 4267–4272 (2007).
22. GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
23. S. M. Lea, *et al.*, Determination of the affinity and kinetic constants for the interaction between the human virus echovirus 11 and its cellular receptor, CD55. *J Biol Chem* **273**, 30443–30447 (1998).
24. I. G. Goodfellow, *et al.*, Inhibition of coxsackie B virus infection by soluble forms of its receptors: binding affinities, altered particle formation, and competition with cellular receptors. *J Virol* **79**, 12016–12024 (2005).
25. C. B. Coyne, J. M. Bergelson, Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. *Cell* **124**, 119–131 (2006).
26. S.-K. Chung, *et al.*, Internalization and trafficking mechanisms of coxsackievirus B3 in HeLa cells. *Virology* **333**, 31–40 (2005).
27. M. M. Davidson, *et al.*, Novel cell lines derived from adult human ventricular cardiomyocytes. *J Mol Cell Cardiol* **39**, 133–147 (2005).
28. M. Shah, *et al.*, Profiling Subcellular Protein Phosphatase Responses to Coxsackievirus B3 Infection of Cardiomyocytes. *Mol Cell Proteomics* **16**, S244–S262 (2017).
29. F. Edfors, *et al.*, Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* **12**, 883 (2016).
30. Y. Liu, *et al.*, Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol* **37**, 314–322 (2019).
31. K. J. D. A. Excoffon, A. Hruska-Hageman, M. Klotz, G. L. Traver, J. Zabner, A role for the PDZ-binding domain of the coxsackie B virus and adenovirus receptor (CAR) in cell adhesion and growth. *J Cell Sci* **117**, 4401–4409 (2004).

32. B. J. Kempf, D. J. Barton, Poliovirus 2A(Pro) increases viral mRNA and polysome stability coordinately in time with cleavage of eIF4G. *J Virol* **82**, 5847–5859 (2008).
33. I. Goodfellow, *et al.*, Identification of a cis-acting replication element within the poliovirus coding region. *J Virol* **74**, 4590–4600 (2000).
34. J. E. Novak, K. Kirkegaard, Coupling between genome translation and replication in an RNA virus. *Genes Dev* **8**, 1726–1737 (1994).
35. A. Cuconati, A. Molla, E. Wimmer, Brefeldin A inhibits cell-free, de novo synthesis of poliovirus. *J Virol* **72**, 6456–6464 (1998).
36. S. Miller, J. Krijnse-Locker, Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**, 363–374 (2008).
37. C. E. Melia, *et al.*, Origins of Enterovirus Replication Organelles Established by Whole-Cell Electron Microscopy. *mBio* **10**, e00951-19 (2019).
38. M. Binder, *et al.*, Replication vesicles are load- and choke-points in the hepatitis C virus lifecycle. *PLoS Pathog* **9**, e1003561 (2013).
39. S. Boersma, *et al.*, Translation and Replication Dynamics of Single RNA Viruses. *Cell* **183**, 1930-1945.e23 (2020).
40. R. W. A. L. Limpens, *et al.*, The transformation of enterovirus replication structures: a three-dimensional study of single- and double-membrane compartments. *mBio* **2**, e00166-11 (2011).
41. M. Bessaud, *et al.*, Development of a Taqman RT-PCR assay for the detection and quantification of negatively stranded RNA of human enteroviruses: evidence for false-priming and improvement by tagged RT-PCR. *J Virol Methods* **153**, 182–189 (2008).
42. S. Singh, *et al.*, In situ 10-cell RNA sequencing in tissue and tumor biopsy samples. *Sci Rep* **9**, 4836 (2019).
43. M. B. Schulte, J. A. Draghi, J. B. Plotkin, R. Andino, Experimentally guided models reveal replication principles that shape the mutation distribution of RNA viruses. *Elife* **4**, e03753 (2015).
44. J. E. Novak, K. Kirkegaard, Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. *J Virol* **65**, 3384–3387 (1991).
45. C. Hohenadl, K. Klingel, J. Mertsching, P. H. Hofschneider, R. Kandolf, Strand-specific detection of enteroviral RNA in myocardial tissue by in situ hybridization. *Mol Cell Probes* **5**, 11–20 (1991).
46. P. E. Tam, R. P. Messner, Molecular mechanisms of coxsackievirus persistence in chronic inflammatory myopathy: viral RNA persists through formation of a double-

- stranded complex without associated genomic mutations or evolution. *J Virol* **73**, 10113–10121 (1999).
47. N. L  v  que, *et al.*, Quantitative genomic and antigenomic enterovirus RNA detection in explanted heart tissue samples from patients with end-stage idiopathic dilated cardiomyopathy. *J Clin Microbiol* **50**, 3378–3380 (2012).
 48. M. B. Schulte, R. Andino, Single-cell analysis uncovers extensive biological noise in poliovirus replication. *J Virol* **88**, 6205–6212 (2014).
 49. W. Xiang, A. Cuconati, D. Hope, K. Kirkegaard, E. Wimmer, Complete protein linkage map of poliovirus P3 proteins: interaction of polymerase 3Dpol with VPg and with genetic variants of 3AB. *J Virol* **72**, 6732–6741 (1998).
 50. Y. Liu, *et al.*, Direct interaction between two viral proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus morphogenesis. *PLoS Pathog* **6**, e1001066 (2010).
 51. P. Jiang, Y. Liu, H.-C. Ma, A. V. Paul, E. Wimmer, Picornavirus morphogenesis. *Microbiol Mol Biol Rev* **78**, 418–437 (2014).
 52. J. E. Darnell, L. Levintow, Poliovirus protein: source of amino acids and time course of synthesis. *J Biol Chem* **235**, 74–77 (1960).
 53. R. Fernandez-Munoz, J. E. Darnell, Structural difference between the 5' termini of viral and cellular mRNA in poliovirus-infected cells: possible basis for the inhibition of host protein synthesis. *J Virol* **18**, 719–726 (1976).
 54. J. M. Lyle, E. Bullitt, K. Bienz, K. Kirkegaard, Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science* **296**, 2218–2222 (2002).
 55. K. R. Gleitsman, R. N. Sengupta, D. Herschlag, Slow molecular recognition by RNA. *RNA* **23**, 1745–1753 (2017).
 56. C. Li, J. C.-Y. Wang, M. W. Taylor, A. Zlotnick, In vitro assembly of an empty picornavirus capsid follows a dodecahedral path. *J Virol* **86**, 13062–13069 (2012).
 57. C. Wang, P. Jiang, C. Sand, A. V. Paul, E. Wimmer, Alanine scanning of poliovirus 2CATPase reveals new genetic evidence that capsid protein/2CATPase interactions are essential for morphogenesis. *J Virol* **86**, 9964–9975 (2012).
 58. C. Zaragoza, *et al.*, Viral protease cleavage of inhibitor of kappaBalpha triggers host cell apoptosis. *Proc Natl Acad Sci U S A* **103**, 19051–19056 (2006).
 59. X. Tan, L. Sun, J. Chen, Z. J. Chen, Detection of Microbial Infections Through Innate Immune Sensing of Nucleic Acids. *Annu Rev Microbiol* **72**, 447–478 (2018).
 60. K. Esser-Nobis, L. D. Hatfield, M. Gale, Spatiotemporal dynamics of innate immune signaling via RIG-I-like receptors. *Proc Natl Acad Sci U S A* **117**, 15778–15788 (2020).

61. R. B. Seth, L. Sun, C.-K. Ea, Z. J. Chen, Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).
62. S. W. Brubaker, A. E. Gauthier, E. W. Mills, N. T. Ingolia, J. C. Kagan, A bicistronic MAVS transcript highlights a class of truncated variants in antiviral immunity. *Cell* **156**, 800–811 (2014).
63. N. Qi, *et al.*, Multiple truncated isoforms of MAVS prevent its spontaneous aggregation in antiviral innate immune signalling. *Nat Commun* **8**, 15676 (2017).
64. Q. Sun, *et al.*, The specific and essential role of MAVS in antiviral innate immune responses. *Immunity* **24**, 633–642 (2006).
65. J. Wu, Z. J. Chen, Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**, 461–488 (2014).
66. S. L. Schwartz, G. L. Conn, RNA regulation of the antiviral protein 2'-5'-oligoadenylate synthetase. *Wiley Interdiscip Rev RNA* **10**, e1534 (2019).
67. A. Pindel, A. Sadler, The role of protein kinase R in the interferon response. *J Interferon Cytokine Res* **31**, 59–70 (2011).
68. A. Rahnefeld, *et al.*, Ubiquitin-like protein ISG15 (interferon-stimulated gene of 15 kDa) in host defense against heart failure in a mouse model of virus-induced cardiomyopathy. *Circulation* **130**, 1589–1600 (2014).
69. Y. Zhang, *et al.*, PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to enhance interferon signaling and control viral infection. *Nat Immunol* **16**, 1215–1227 (2015).
70. D. Franco, *et al.*, Stimulation of poliovirus RNA synthesis and virus maturation in a HeLa cell-free in vitro translation-RNA replication system by viral protein 3CDpro. *Virology* **2**, 86 (2005).
71. C. Zitzmann, *et al.*, A Coupled Mathematical Model of the Intracellular Replication of Dengue Virus and the Host Cell Immune Response to Infection. *Front Microbiol* **11**, 725 (2020).
72. M. R. Patel, Y.-M. Loo, S. M. Horner, M. Gale, H. S. Malik, Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol* **10**, e1001282 (2012).
73. N. Blom, J. Hansen, D. Blaas, S. Brunak, Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci* **5**, 2203–2216 (1996).
74. S. C. A. Nielsen, *et al.*, Probable transmission of coxsackie B3 virus from human to chimpanzee, Denmark. *Emerg Infect Dis* **18**, 1163–1165 (2012).

75. M. Takada, Y. Yamane, T. Nagashima, T. Asahara, Y. Owada, Various antibody holding status against viruses in *Cercopithecus aethiops* and *Macaca irus*. 3. The presence of antibodies to coxsackie, rubella, Japanese encephalitis, influenza and parainfluenza viruses. *Kitasato Arch Exp Med* **41**, 135–151 (1968).
76. F. Deinhardt, A. Holmes, J. Devine, J. Deinhardt, Marmosets as Laboratory Animals .4. Microbiology of Laboratory Kept Marmosets. *Lab Anim Care* **17**, 48–70 (1967).
77. E. Meylan, *et al.*, Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
78. L. Phan, *et al.*, ALFA: Allele Frequency Aggregator. *National Center for Biotechnology Information* (2020).
79. J. Pothlichet, *et al.*, A loss-of-function variant of the antiviral molecule MAVS is associated with a subset of systemic lupus patients. *EMBO Mol Med* **3**, 142–152 (2011).
80. F. Xing, *et al.*, Alteration of Antiviral Signalling by Single Nucleotide Polymorphisms (SNPs) of Mitochondrial Antiviral Signalling Protein (MAVS). *PLoS One* **11**, e0151173 (2016).
81. D. Schweinoch, P. Bachmann, D. Clausnitzer, M. Binder, L. Kaderali, Mechanistic modeling explains the dsRNA length-dependent activation of the RIG-I mediated immune response. *J Theor Biol* **500**, 110336 (2020).
82. F. Hou, *et al.*, MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146**, 448–461 (2011).
83. B. Wu, *et al.*, Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell* **152**, 276–289 (2013).
84. B. Wu, *et al.*, Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Mol Cell* **55**, 511–523 (2014).
85. J. F. Enders, T. H. Weller, F. C. Robbins, Cultivation of the Lansing Strain of Poliomyelitis Virus in Cultures of Various Human Embryonic Tissues. *Science* **109**, 85–87 (1949).
86. B. M. McDermott, A. H. Rux, R. J. Eisenberg, G. H. Cohen, V. R. Racaniello, Two distinct binding affinities of poliovirus for its cellular receptor. *J Biol Chem* **275**, 23089–23096 (2000).
87. C. M. Kunin, Virus-tissue union and the pathogenesis of enterovirus infections. *J Immunol* **88**, 556–569 (1962).
88. Z. Jin, J. Deval, K. A. Johnson, D. C. Swinney, Characterization of the elongation complex of dengue virus RNA polymerase: assembly, kinetics of nucleotide incorporation, and fidelity. *J Biol Chem* **286**, 2067–2077 (2011).

89. B. Brown, M. S. Oberste, K. Maher, M. A. Pallansch, Complete genomic sequencing shows that polioviruses and members of human enterovirus species C are closely related in the noncapsid coding region. *J Virol* **77**, 8973–8984 (2003).
90. J. Herold, R. Andino, Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell* **7**, 581–591 (2001).
91. Y. Yang, *et al.*, Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc Natl Acad Sci U S A* **104**, 7253–7258 (2007).
92. M.-S. Hwang, *et al.*, MAVS polymers smaller than 80 nm induce mitochondrial membrane remodeling and interferon signaling. *FEBS J* **286**, 1543–1560 (2019).
93. K. A. Janes, D. A. Lauffenburger, Models of signalling networks - what cell biologists can gain from them and give to them. *J Cell Sci* **126**, 1913–1921 (2013).
94. W. M. Koh, *et al.*, The Epidemiology of Hand, Foot and Mouth Disease in Asia: A Systematic Review and Analysis. *Pediatr Infect Dis J* **35**, e285-300 (2016).
95. D. Becker, *et al.*, Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments. *Nat Commun* **7**, 12761 (2016).
96. F. G. Hayden, *et al.*, Phase II, randomized, double-blind, placebo-controlled studies of rupintrivir nasal spray 2-percent suspension for prevention and treatment of experimentally induced rhinovirus colds in healthy volunteers. *Antimicrob Agents Chemother* **47**, 3907–3916 (2003).
97. Y. Kim, *et al.*, Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J Virol* **86**, 11754–11762 (2012).
98. X. Li, *et al.*, Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* **6**, eabb9153 (2020).
99. X. Yang, *et al.*, A public genome-scale lentiviral expression library of human ORFs. *Nat Methods* **8**, 659–661 (2011).
100. R. Kandolf, P. H. Hofschneider, Molecular cloning of the genome of a cardiotropic Coxsackie B3 virus: full-length reverse-transcribed recombinant cDNA generates infectious virus in mammalian cells. *Proc Natl Acad Sci U S A* **82**, 4818–4822 (1985).
101. K. J. Jensen, *et al.*, An ERK-p38 subnetwork coordinates host cell apoptosis and necrosis during coxsackievirus B3 infection. *Cell Host Microbe* **13**, 67–76 (2013).
102. K. A. Janes, An analysis of critical factors for quantitative immunoblotting. *Sci Signal* **8**, rs2 (2015).
103. K.-S. Kim, *et al.*, 5'-Terminal deletions occur in coxsackievirus B3 during replication in murine hearts and cardiac myocyte cultures and correlate with encapsidation of negative-strand viral RNA. *J Virol* **79**, 7024–7041 (2005).

104. K. Miller-Jensen, K. A. Janes, J. S. Brugge, D. A. Lauffenburger, Common effector processing mediates cell-specific responses to stimuli. *Nature* **448**, 604–608 (2007).
105. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
106. R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750-753 (2010).
107. Q. M. Wang, *et al.*, Enzymatic characterization of refolded human rhinovirus type 14 2A protease expressed in *Escherichia coli*. *J Virol* **72**, 1683–1687 (1998).
108. L. Shang, *et al.*, Biochemical characterization of recombinant Enterovirus 71 3C protease with fluorogenic model peptide substrates and development of a biochemical assay. *Antimicrob Agents Chemother* **59**, 1827–1836 (2015).
109. A. J. Dorner, *et al.*, In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate. *J Virol* **50**, 507–514 (1984).
110. J. J. Arnold, C. E. Cameron, Poliovirus RNA-dependent RNA polymerase (3D(pol)). Assembly of stable, elongation-competent complexes by using a symmetrical primer-template substrate (sym/sub). *J Biol Chem* **275**, 5329–5336 (2000).
111. R. R. Regoes, S. Crotty, R. Antia, M. M. Tanaka, Optimal replication of poliovirus within cells. *Am Nat* **165**, 364–373 (2005).
112. D. Endres, A. Zlotnick, Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* **83**, 1217–1230 (2002).
113. A. Zlotnick, Are weak protein-protein interactions the general rule in capsid assembly? *Virology* **315**, 269–274 (2003).
114. C. G. Lopez, O. Saldanha, K. Huber, S. Köster, Lateral association and elongation of vimentin intermediate filament proteins: A time-resolved light-scattering study. *Proc Natl Acad Sci U S A* **113**, 11152–11157 (2016).

Chapter 3:

Simulating Coxsackievirus B3 infection with an accessible computational model of its complete kinetics

Cameron D. Griffiths,^{1,3,4,*} Andrew J. Sweatt,^{1,3} and Kevin A. Janes^{1,2,5,**}

¹Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA.

²Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA.

³These authors contributed equally.

⁴Technical contact

⁵Lead contact

*Correspondence: cg3gh@virginia.edu (C.D.G.)

**Correspondence: kjan@virginia.edu (K.A.J.)

This work has been published in *STAR Protocols* and can be found at <https://doi.org/10.1016/j.xpro.2021.100940>.

Data and code availability

The graphical user interface used during this protocol is available at <https://github.com/JanesLab/CompleteKinetics-CVB3-GUI>.

3.1 Abstract

We describe how to use a publicly available computational model for coxsackievirus B3 (CVB3) infection that we recast as a graphical user interface (GUI). The GUI-based implementation enables non-computationalists to incorporate systems-biology modeling into their research and teaching. The model simulates the full life cycle of CVB3, including the host antiviral response, and includes 44 alterable parameters. The model simplifies some viral life cycle processes to improve interpretability and utility when performing *in silico* experiments.

For complete details on the use and execution of this protocol, please refer to (2).

3.2 Before you begin

3.2.1 Background and implementation

Computational models are essential tools for hypothesis generation and testing in systems biology (1). However, most computational models have limited documentation and require knowledge of structured programming to operate. To overcome this barrier for non-computationalists, we developed a graphical user interface (GUI) version of a publicly available model of coxsackievirus B3 (CVB3) infection (2). CVB3 is a member of the Picornavirus family of non-enveloped, positive-strand RNA viruses. Picornaviruses are a well-studied group of viruses, with fully delineated life cycles from viral docking and entry to cell lysis (3). After entry into the cytoplasm, the compact picornaviral genome is rapidly translated as a single polypeptide, yielding an equal proportion of viral proteins. From a computational modeling perspective, the abundance of quantitative information and one-to-one protein stoichiometry make it plausible to define a complete kinetic model of CVB3 infection. Complete kinetic models maintain stoichiometric constraints and seek to explain how reactants are converted to products from start to finish (2). The GUI maintains the predictive and explanatory power of the CVB3 complete kinetic model and adds user-friendliness for non-computationalists.

The model encodes the entire CVB3 lifecycle as a system of 54 differential equations (2). To understand many of the input parameters and the model output, it is important to be familiar with the CVB3 life cycle. The model breaks down the viral life cycle into three main modules: delivery, replication, and encapsidation. The modules are overlaid with a set of negative feedbacks that capture the effects of several interferon-stimulated genes (ISGs) that approximate the innate immune response of the host cell.

Delivery module

CVB3 uses a multi-step process to enter a host cell. The virus first binds to decay-accelerating factor (DAF), a protein receptor on the cell surface (4). Then, CVB3 translocates to tight junctions and binds to a second receptor called coxsackievirus and adenovirus receptor (CAR) (5). After binding CAR, CVB3 is internalized via endocytosis, and the positive-sense RNA genome of the virus is delivered into the cytoplasm (6). In addition to live particles, stock preparations of CVB3 also contain many defective RNA-filled particles that are unable to replicate. For CVB3, the live-to-defective particle ratio is approximately 1:800, depending on the viral stock (2). In the model, defective particles follow a delivery process similar to live particles, but the defective RNA is not translated or replicated.

Replication module

Once in the cytoplasm, the CVB3 genome is translated by host ribosomes into a single polyprotein in a cap-independent manner (3). The CVB3 genome is simultaneously bound by multiple ribosomes (polysome), forming a translation complex (7). After translation, the polyprotein matures through a series of cleavage events into individual viral proteins, including the viral RNA-dependent RNA-polymerase ($3D^{pol}$), two proteases ($2A^{pro}$ and $3C^{pro}$), an ATPase ($2C^{ATPase}$), and structural capsid proteins (VP1–4) (3). In parallel, the positive-sense CVB3 genome is replicated into full-length negative-sense genomes by the viral $3D^{pol}$. The negative-sense genomes are subsequently replicated by $3D^{pol}$ to generate more positive-sense genomes (3). During replication, CVB3 hydrophobic proteins (including $2C^{ATPase}$) recruit intracellular membranes to form viral replication organelles (VROs) (8, 9). The VROs serve as a platform for replication, concentrating viral proteins and accelerating replication.

Encapsidation module

To create new viral particles, CVB3 capsid proteins self-assemble with positive-sense viral RNA. First, three capsid proteins (VP0 [the precursor for VP2 and VP4], VP1, and VP3) bind to form a capsid protomer (10). In the model, a protomer contributes immediately to one fifth of a self-assembled pentamer. Twelve pentamers are assumed to bind sequentially to form a mature viral particle. During pentamer recruitment, a positive-sense viral RNA strand may or may not be bound to a pentamer. If viral RNA is bound when the twelfth pentamer is added in the model, a filled (infectious) virion is formed. Otherwise, an empty (non-infectious) virion is formed. At any point during capsid formation, pentamers or viral RNA may unbind, but once fully assembled, the model assumes that mature virions

do not disassemble into individual pentamers.

Feedbacks involving the innate immune response

Host cells have mechanisms to detect infecting viruses and mount an antiviral response through a series of events that shut down transcription and translation. Reciprocally, viruses encode countermeasures to block the antiviral response. For CVB3, melanoma differentiation-associated protein 5 (MDA5) recognizes double-stranded RNA formed during CVB3 replication (11, 12). MDA5 bound to double-stranded RNA signals through mitochondrial antiviral-signaling protein (MAVS) to stimulate type I interferon (IFN) production (13, 14). Type I IFN is then secreted, stimulating the infected cell (as well as nearby cells) to induce ISG expression (15). CVB3 proteases 2A^{pro} and 3C^{pro} antagonize the innate immune response by cleaving host-cell proteins, including MAVS (16, 17). Furthermore, 2A^{pro} cleaves eukaryotic translation initiation factor 4G (eIF4G) (18), which prevents translation of host mRNAs but does not affect translation of the CVB3 genome.

The model encodes the innate immune response as negative feedbacks that inhibit translation, inactivate proteases, and accelerate RNA degradation. Viral antagonism is encoded as a counteracting negative feedback that decreases the host-cell ability to initiate an immune response.

Modeling implementation

The published model is publicly available as a MATLAB package at <https://github.com/JanesLab/CompleteKinetics-CVB3>. To address the drawback that the published model requires structured programming knowledge and commercial software, we built a standalone, open-source version of the model that runs simulations behind a GUI. The protocol below describes the installation and use of the CVB3 complete kinetics GUI; more-advanced users are referred to the original research publication describing the MATLAB package (2).

3.2.2 Install the graphical user interface (GUI)

Timing: 5 – 30 min

The CVB3 complete kinetics GUI requires installation of MATLAB Runtime version 9.10 (R2021a; 724 MB download size). The CVB3 GUI and MATLAB Runtime software can be installed automatically or manually on Windows or MacOS operating systems. The automatic installation uses an installation wizard to install both MATLAB Runtime (if needed) and the GUI. The manual installation is provided as an alternative, where MATLAB

Runtime is separately installed and the GUI is run directly from the .exe (Windows) or .app (MacOS) file. The files for the following download steps can be found on the JanesLab Github site (see **3.3 Key resources table**).

Note: For Linux users, we compiled the GUI on a Centos distribution and made the compiled application available on the JanesLab Github site (see **3.3 Key resources table**). Installing and running the GUI on Linux requires knowledge of command line-based scripting, which is beyond the scope of this protocol. The instructions for installing and running the GUI can be found in the readme file provided in the Linux folder on the JanesLab Github site.

1. Windows Automatic Installation Instructions

- a. Download CVB3_AppInstaller_web.exe, then double click to run the program.
- b. Click “Yes” to allow this app from an unknown publisher to make changes to your device.
- c. Click “Next >” to advance past the installation start page.
- d. If desired, check “Add a shortcut to the desktop”.
- e. Click “Next >” to confirm the CVB3 GUI installation folder selection.
- f. At this point, the installer checks if MATLAB Runtime is installed. If not, perform the following steps:
 - i. Click “Next >” to confirm the MATLAB Runtime installation folder selection.
 - ii. Check “Yes” then click “Next >” to agree to the MATLAB Runtime user agreement.
- g. Click “Install >” to install MATLAB Runtime (if needed) and the CVB3 GUI.
- h. Click “Finish” to close the installer.
- i. Run the CVB3 GUI like any other program.

2. Windows Manual Installation Instructions

- a. Verify that version 9.10 (R2021a) of MATLAB Runtime is installed. If installed, MATLAB Runtime can be found in the “Add or remove programs” list, which can be found by searching for “Add or remove programs” in the computer’s search bar.
 - i. If MATLAB Runtime is not installed, we advise following the “Windows Automatic Installation Instructions” to download MATLAB Runtime

and the GUI simultaneously. Otherwise, download and install the appropriate version of MATLAB Runtime for R2021a from the following link on the MathWorks website: <https://www.mathworks.com/products/compiler/mcr/index.html>.

- b. Download the CVB3_GUI_Manual_Install folder to a desired location on the hard drive.
- c. Run the CVB3 GUI by double-clicking the CVB3_GUI.exe icon.

3. MacOS Automatic Installation Instructions

- a. Download CVB3_AppInstaller_web.app.zip.
- b. Right click on the file and select “Open” to open the installer.
 - i. The following message may appear: “macOS cannot verify the developer of “CVB3_AppInstaller_web” Are you sure you want to open it?”. Click “Open” to proceed with the installation.
 - ii. If needed, enter the administrator User Name and Password and click “OK”.
- c. Click “Next >” to advance past the installation start page.
- d. Choose an installation folder and click “Next >” to confirm the CVB3 GUI installation folder selection.
- e. At this point, the installer checks if MATLAB Runtime is installed. If not, perform the following steps:
 - i. Choose an installation folder for MATLAB Runtime and click “Next >” to confirm the MATLAB Runtime installation folder selection.

Note: MATLAB Runtime must be installed in a folder named “MATLAB” in the Applications directory (Figure 3.1). If the application is installed in another directory, copy it to the Applications directory.

- ii. Check “Yes” then click “Next >” to agree to the MATLAB Runtime user agreement.
- f. Click “Install >” to install MATLAB Runtime (if needed) and the CVB3 GUI.
- g. Click “Finish” to close the installer.
- h. Open the Janes_Lab folder (in the install location from step d—usually “Applications”).
- i. Open the CVB3_GUI folder.
- j. Open the application folder.

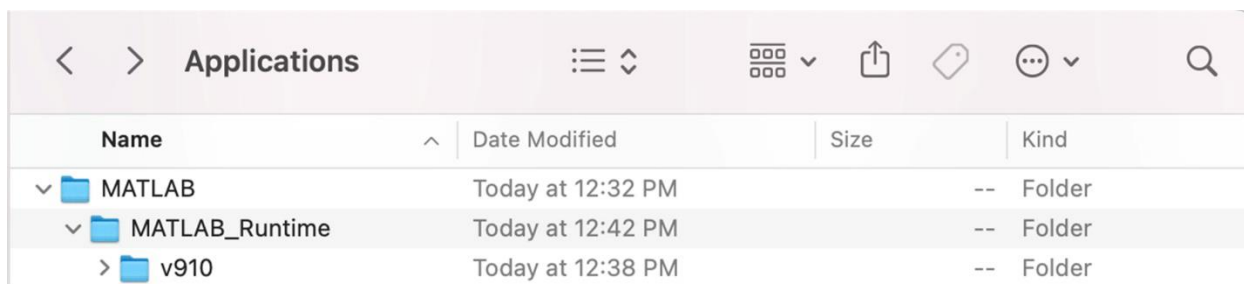
- k. If desired, make an alias of CVB3_GUI.app for easier access.
- l. Run the CVB3 GUI by double-clicking the CVB3_GUI.app icon.

4. MacOS Manual Installation Instructions

- a. Verify that version 9.10 (R2021a) of MATLAB Runtime is installed. If installed, MATLAB Runtime can be found in the Applications directory.
 - i. If MATLAB Runtime is not installed, we advise following the “MacOS Automatic Installation Instructions” to download MATLAB Runtime and the GUI simultaneously. Otherwise, download and install the appropriate version of MATLAB Runtime for R2021a from the following link on the MathWorks website: <https://www.mathworks.com/products/compiler/mcr/index.html>.

Note: MATLAB Runtime must be installed in a folder named “MATLAB” in the Applications directory (Figure 3.1). If the application is installed in another directory, copy it to the Applications directory.

- b. Download the CVB3_GUI_Manual_Install folder to a desired location on the hard drive.
- c. Open the CVB3_GUI_Manual_Install folder.
- d. Double click on the CVB3_GUI.app.zip icon.
- e. Run the CVB3 GUI by right clicking the CVB3_GUI icon and selecting “Open”.
 - i. The following message may appear: “macOS cannot verify the developer of “CVB3_AppInstaller_web”. Are you sure you want to open it?”. Click “Open” to proceed with running the CVB3 GUI.
 - ii. If needed, enter the administrator User Name and Password and click “OK”.



Applications			
Name	Date Modified	Size	Kind
▼ MATLAB	Today at 12:32 PM	--	Folder
▼ MATLAB_Runtime	Today at 12:42 PM	--	Folder
▶ v910	Today at 12:38 PM	--	Folder

Figure 3.1. MacOS folder setup for MATLAB Runtime.

For the CVB3 model to run correctly on MacOS, MATLAB Runtime must be installed in the MATLAB folder within the Applications directory. The v910 folder within “MATLAB Runtime” is set up automatically when installing MATLAB Runtime.

3.3 Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
CVB3 complete kinetics computational model	Lopacinski et al., 2021	https://github.com/JanesLab/CompleteKinetics-CVB3
CVB3 complete kinetics GUI	This paper	https://github.com/JanesLab/CompleteKinetics-CVB3-GUI
MATLAB R2021a	MathWorks	https://www.mathworks.com/products/matlab.html
MATLAB Runtime version 9.10 (R2021a)	MathWorks	https://www.mathworks.com/products/compiler/matlab-runtime.html

3.4 Step-by-step method details

3.4.1 Basic operations

Timing: 1 – 2 minutes per simulation

(Simulation time may vary with computer processing speed and non-default parameter sets.)

This protocol describes the basic model functions to provide a starting point for further simulations. The program runs simulations behind the GUI and stores the concentration of 59 output species as a function of time (e.g., virions, total single-stranded RNA, positive-sense RNA, viral protein, and interferon stimulated proteins). The user selects one species to be plotted on the main GUI figure. When multiple simulations are run sequentially, the data from previous simulations are saved until one of three things occurs: 1) the Clear Data/Plot button (General tab) is pressed, 2) the simulation mode is switched (Options tab), or 3) a sensitivity analysis is enabled or disabled (Options tab). All simulations that have not been cleared are referred to as active simulations.

To demonstrate basic operations, this protocol begins with a default infection for 16 hours at a multiplicity of infection (MOI) of 10, with all parameters set to literature or experimental estimates (2). At the end of this outline, the reader should be able to simulate infections, change key parameters, and export results. Descriptions of all buttons on the main GUI page are found in Table 3.1.

1. To begin the default simulation, click Start. The median virion concentration over time is automatically plotted along with the upper quantile (default: 95%) and lower quantile (default: 5%) (Figure 3.2A).

Note: The upper and lower quantiles are plotted because the CVB3 complete kinetics GUI performs multiple runs—determined by Number of runs (Options tab)—with slightly different parameter sets for a single simulation. See the **Quantification and statistical analysis** section for details on the variation between runs.

2. To infect with half as much virus, change the number in the MOI box to 5 and click Start. The new simulation is overlaid on the previous simulation in a new color (Figure 3.2B, blue).

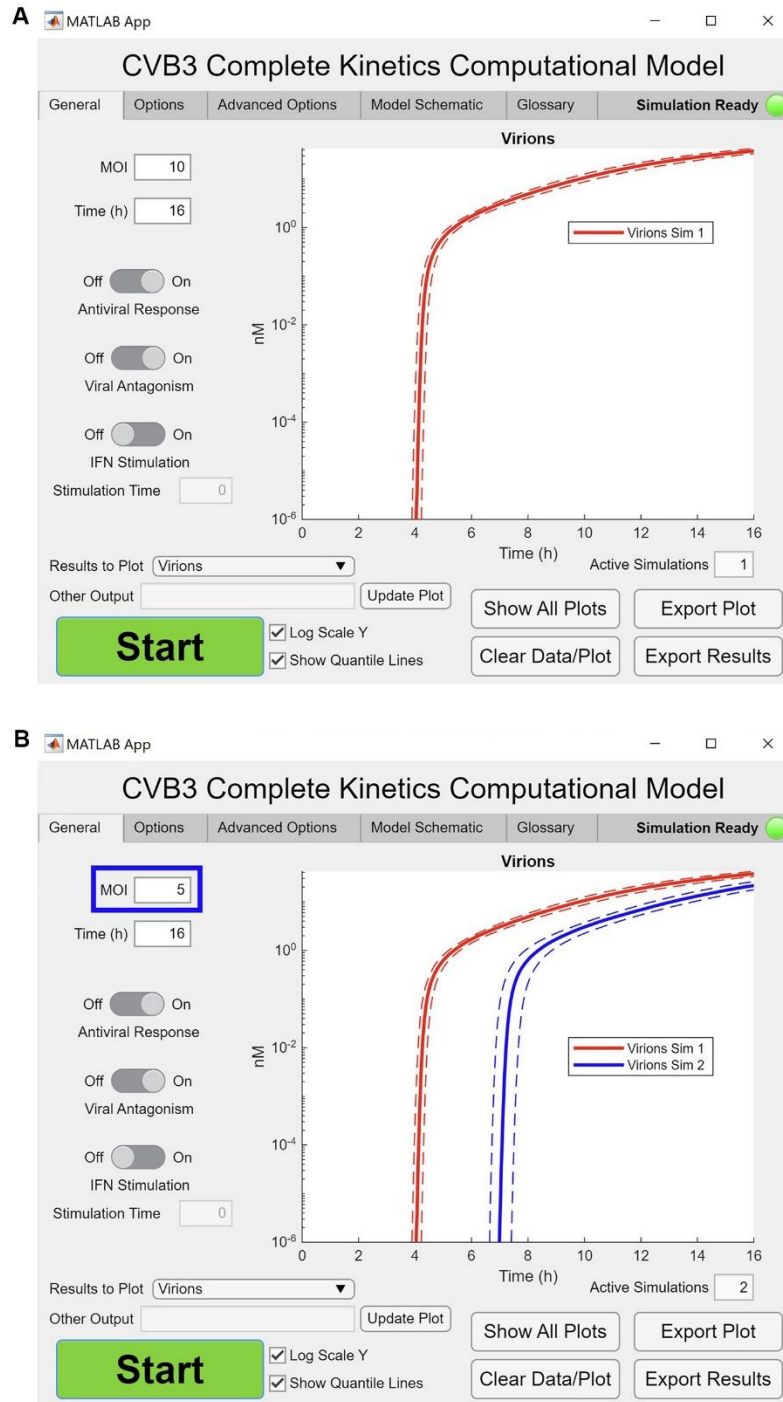


Figure 3.2. Demonstration of a single simulation and sequential simulations.

(A) The graphical output is the result of running the model with default settings. (B) The graphical output of a second simulation (blue) using MOI 5 is overlaid on the first simulation (red). The blue box highlights changes made for the simulation. The solid lines show the median number of virions over time, and the dotted lines show the upper (95%) and lower (5%) quantiles.

Note: Ten unique colors are pre-loaded, and the colors will repeat if more simulations are overlaid.

3. To examine the effect of the host antiviral response, toggle the Antiviral Response switch off and click Start. For simulations at an MOI of 5, the median virion concentration is increased by toggling off the antiviral response (Figure 3.3A, green compared to blue). Next, change the Results to Plot dropdown to Interferon Stimulated Proteins and uncheck the Log Scale Y button. The lack of interferon stimulated proteins confirms that the antiviral response is disabled (Figure 3.3B, green).
4. The model can also simulate the effect of adding exogenous IFN. To perform a simulation with exogenous IFN added 6 hours after infection, toggle the IFN Stimulation switch on, and enter “6” in the Stimulation Time field. Next, toggle the Antiviral Response switch back on and click Start. To plot +ssRNA instead of virion concentration, change the Results to Plot dropdown to Total +ssRNA and check the Log Scale Y button. Uncheck the Show Quantile Lines button to reduce crowding and make the plot easier to interpret (Figure 3.4, cyan compared to blue).

Note: Entering a negative value in the Stimulation Time field simulates IFN pretreatment of cells prior to infection.

Note: The Antiviral Response switch must be turned on for the exogenous IFN to stimulate an antiviral response. Otherwise, the model does not incorporate the exogenous IFN. The model simulation is the same with and without exogenous IFN if the Antiviral Response is off.

5. To save the main figure plot as it is currently displayed, press the Export Plot button. To export another plot with a different species-of-interest, change the Results to Plot dropdown then press the Export Plot button again. Press the Export Results button to save all of the underlying data. Export Results saves 5 tables per active simulation (Median, Mean, Upper Quantile, Lower Quantile, and Run Parameters).

Note: The Run Parameters table is used as a reference for the settings that went into each simulation.

Note: To view many species without changing Results to Plot, press the Show All Plots button. Show All Plots will plot all 59 output species grouped into 7 windows per simulation rather than overlaying the data from multiple simulations. For example, clicking Show All Plots after two simulations will result in 14 windows. These plots can also be saved.

6. In addition to the parameters modified in the above steps, all other parameters in the General tab (Table 3.1), Options tab (Table 3.2), or Advanced Options tab (Table 3.3) can also be edited. Descriptions of all fields and their default values are listed in Tables 3.1–3.

Note: All parameters can be reset to defaults values by clicking the Reset all Parameters to Default Values button in the Advanced Options tab. Resetting to default values does not clear the plot, remove active simulations, change simulation mode, or enable/disable sensitivity analysis.

Note: Simulations that result in a median virion concentration greater than 326 nM will be truncated at the time point that 326 nM is reached. See the **Quantification and statistical analysis** for the rationale and calculation of the threshold.

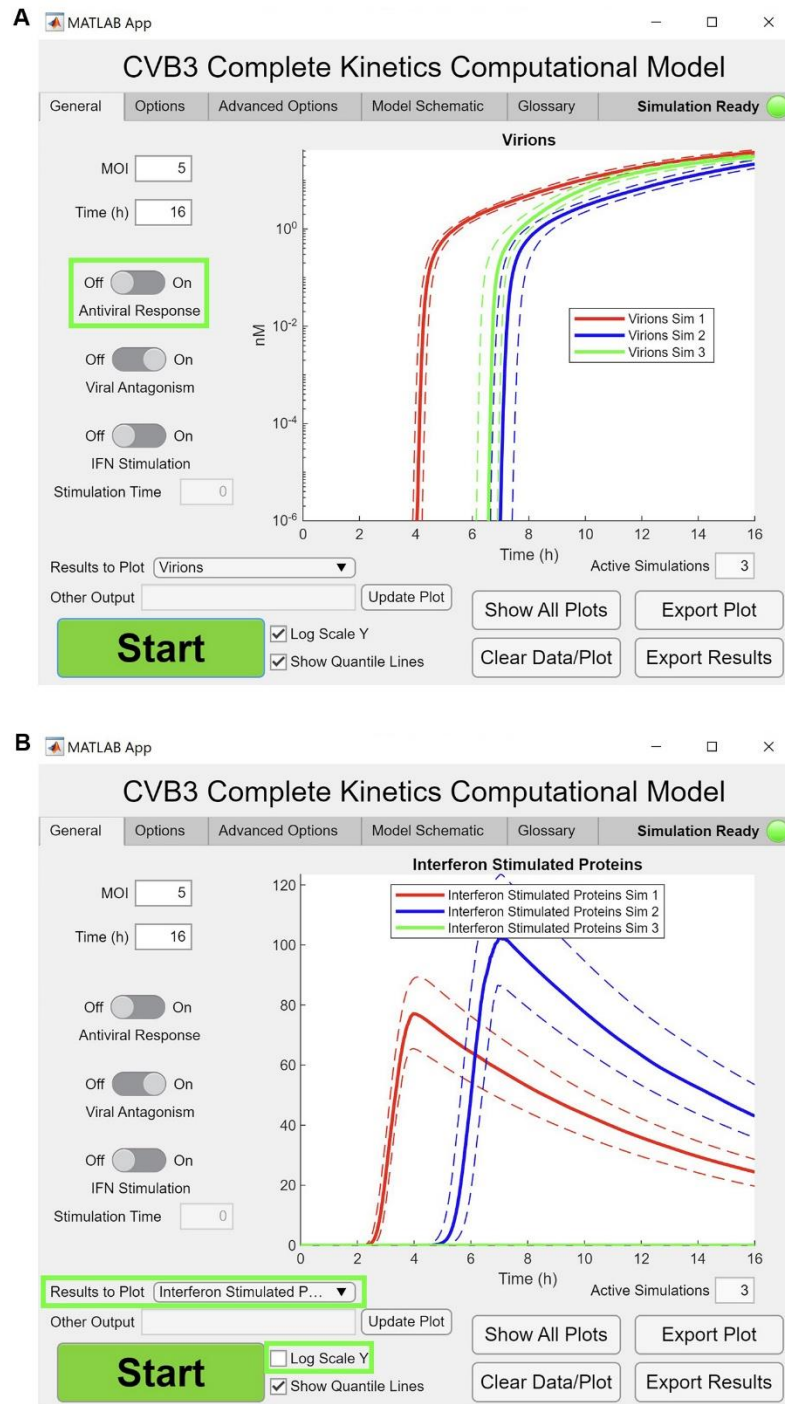


Figure 3.3. Turning off the antiviral response and plotting a different species.

The graphical output is the result of running a third simulation (green) after the simulations shown in Figure 2 (red and blue). The third simulation was run at an MOI of 5 with the cellular antiviral response turned off. **(A)** Virion production increases without an antiviral response. **(B)** No interferon stimulated proteins are produced when the antiviral response is turned off. The green box highlights changes made for the simulation.

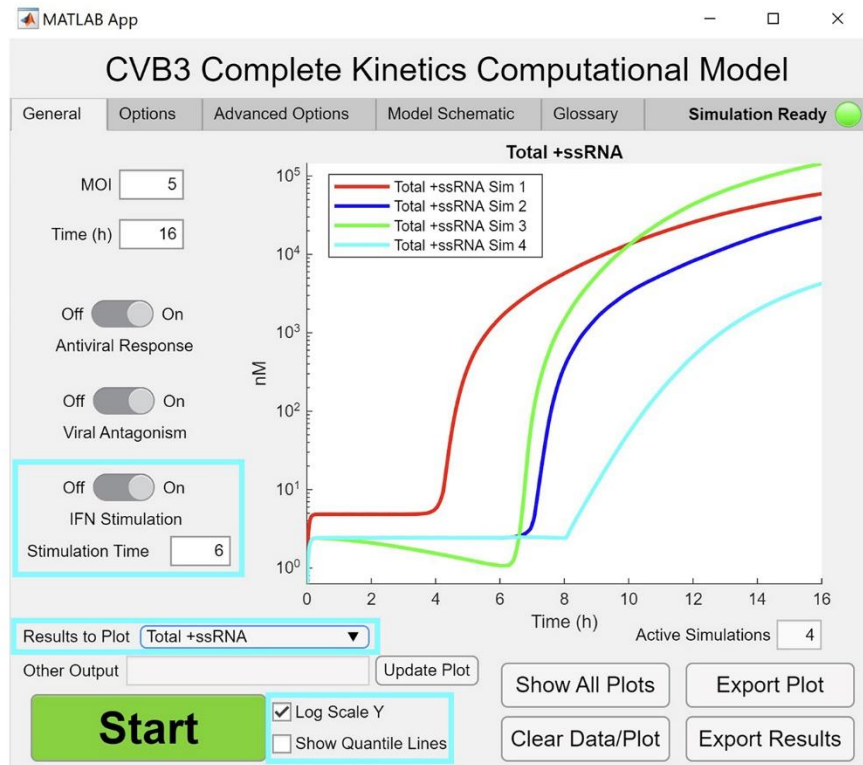


Figure 3.4. Adding exogenous IFN stimulation.

The graphical output is the result of running a fourth simulation (cyan) after the simulation shown in Figure 3.3. The fourth simulation was run at an MOI of 5 with exogenous IFN stimulation starting at 6 hours post infection and the host antiviral response on. The plot shows total positive-sense single-stranded RNA (+ssRNA) on a logscale y-axis with the quantile lines turned off. The cyan boxes highlight changes made for the simulation.

3.4.2 Population mode

Timing: 1 – 2 minutes per simulation

(Simulation time may vary with computer processing speed and non-default parameter sets.)

The default simulation mode is to model the infection of a single cell with the number of virions specified in the MOI box. The model is also able to simulate infection of an entire population of cells with the discrete number of virions for each run taken from a Poisson distribution around a mean specified by the MOI, as occurs in an actual infection (19). In population mode, each run involves a different number of virions; the mean and median for output species are plotted as the population-wide average without quantile lines.

7. Click Reset all Parameters to Default Values under the Advanced Options tab to undo all changes made thus far and reset the model. Under the Options tab, turn the Simulation Mode knob from Single Cell to Population, which will clear all active simulations. Under the General tab, click Start. Both the median and the mean virion concentration are plotted without quantile lines (Figure 3.5).

Note: Quantile lines are disabled in population mode. No other GUI functionality is lost. Like in single cell mode, all other model parameters can be modified, and results can be exported.

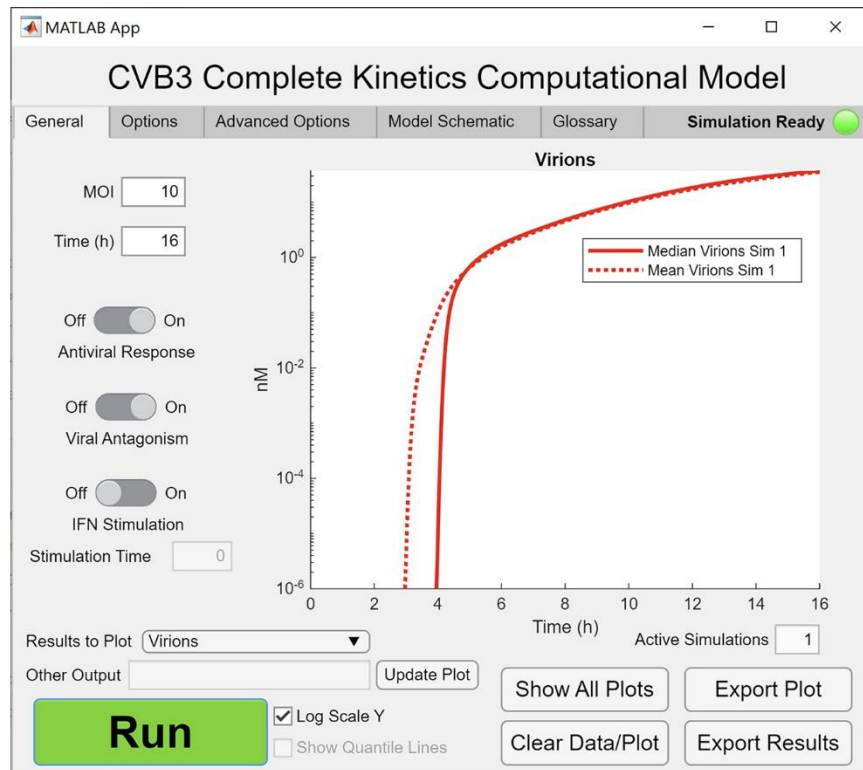


Figure 3.5. A simulation in population mode.

The graphical output is the result of running a simulation in Population mode instead of Single Cell mode. In Population mode, the number of viral particles in each run is randomly selected from a Poisson distribution with the mean set at the input MOI. The plot shows the mean (dotted line) and the median (solid line) virions over time.

3.4.3 Sensitivity analysis

Timing: 2 – 4 minutes

(Simulation time may vary with computer processing speed and non-default parameter sets.)

The CVB3 computational model can run a sensitivity analysis. The sensitivity analysis shows changes in a user-defined output species when parameters are individually scaled up or down. Each parameter is scaled by the value in the Scaling factor box to a maximum/minimum set by the Max orders of magnitude to scale box. The coefficient of variation among runs is set to 0 so that the parameter being scaled is the only change between iterations. As a result, quantile lines are also disabled. The sensitivity analysis output is displayed as a heatmap showing the fold change in the output species-of-interest at the simulation endpoint.

8. Click Reset all Parameters to Default Values under the Advanced Options tab to reset the model. Under the Options tab, check the Sensitivity analysis box, which will clear all active simulations. Under the General tab, click Start. The default sensitivity analysis shows changes in positive-sense CVB3 RNA (+ssRNA) (Figure 3.6).

Note: A sensitivity analysis can be run in either single cell mode (as in Figure 3.6) or population mode.

Note: The sensitivity analysis output can either be a single species or the sum of any number of output species. Multiple species can be selected by selecting Other in the Output species dropdown menu and entering the output species to sum in the Other output box, separated by commas (no spaces).

Note: Similar to running the model with the sensitivity analysis disabled, any model parameter can be modified, and the results can be exported. An additional table containing the results of the sensitivity analysis will also be exported.

Note: In contrast to running the model with the sensitivity analysis disabled, virion concentrations exceeding 326 nM will not be truncated.

Comprehensive Sensitivity Analysis: Log2 Fold Changes in Positive Strand RNA Concentration at 16.0 hpi

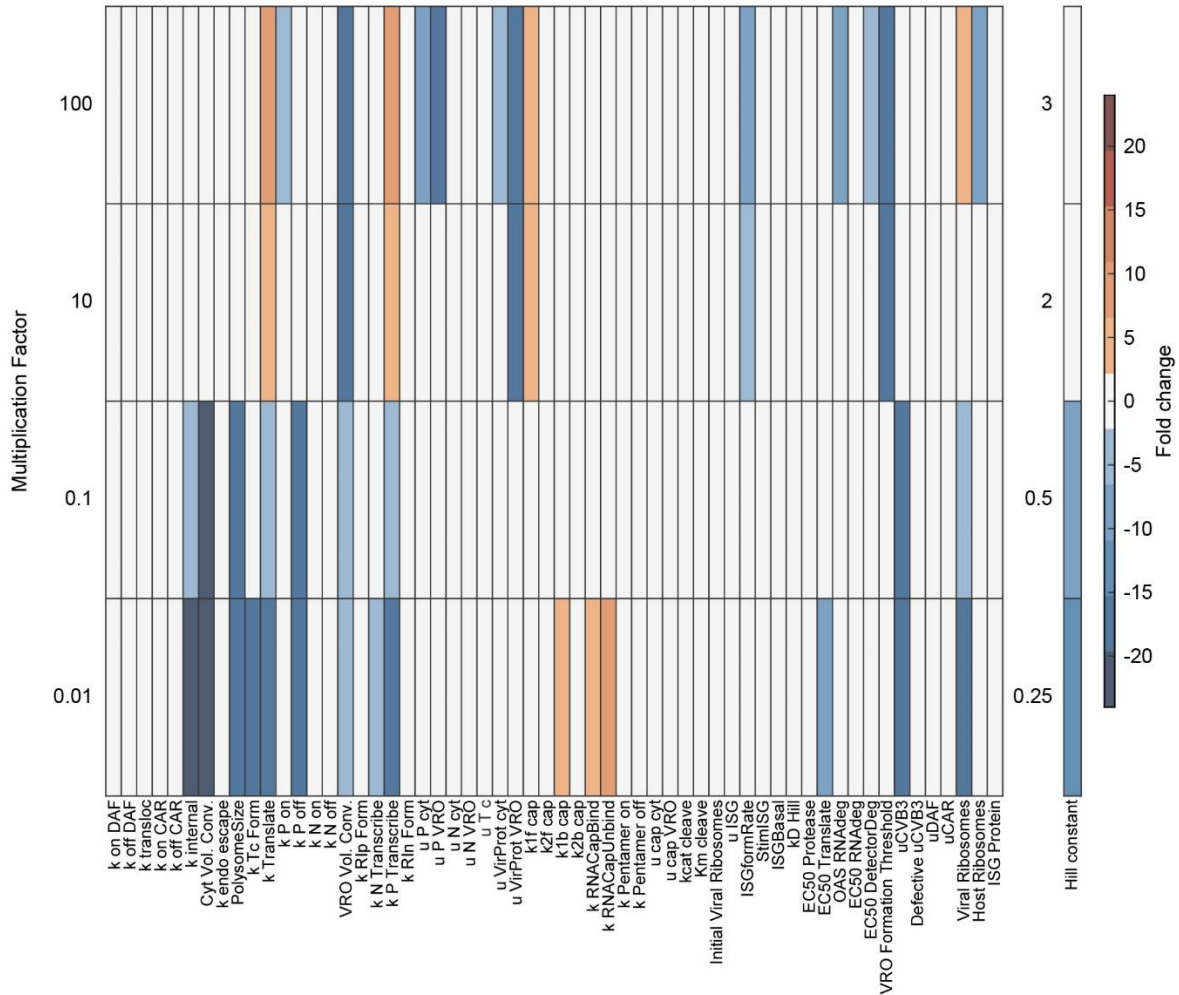


Figure 3.6. Performing a single-parameter sensitivity analysis.

The graphical output is the result of running a sensitivity analysis for positive-strand RNA concentration at 16 hours post infection. On the x-axis are parameters that were individually scaled up or down by the indicated multiplication factor on the y-axis relative to the base parameter set. The Hill constant was varied multiplicatively below one and additively above one, where 2 = base Hill coefficient + 1. Orange rectangles indicate parameter values that increased positive-strand RNA concentration compared to default, whereas blue rectangles indicate decreased positive-strand RNA concentration. The sensitivity analysis plot was processed in Adobe Illustrator.

3.5 Expected outcomes

The CVB3 complete kinetics computational model reproduces the entire life cycle of CVB3. Specific viral life cycle events can be observed by examining different model output species. To illustrate, a simulation was run using default parameters, with the time set to 6 hours (Figure 3.7). CVB3 binding occurs within the first hour, as reflected by the disappearance of unbound CVB3 (Figure 3.7A). However, full viral entry does not finish until shortly after 3 hours post infection, when +ssRNA is no longer present in endosomes (Figure 3.7B). Next, the viral replication organelle (VRO) forms (Figure 3.7C, arrow), which concentrates viral proteins and viral RNA. The concentrating effect rapidly accelerates the formation of ribosome-bound +ssRNA viral translation complexes (Figure 3.7D), viral protein (Figure 3.7E), and +ssRNA (Figure 3.7F) (2). The accumulation of viral proteins initiates the self-assembly of virions. During CVB3 self-assembly, capsid pentamers come together sequentially, along with +ssRNA. When the twelfth pentamer joins, the capsid closes, forming an intact virion. Accordingly, two pentamers bound to +ssRNA is observed before five pentamers bound to +ssRNA, which occurs before full virions are produced (Figure 3.7G–I).

The CVB3 complete kinetics GUI may be used to perform experiments *in silico*. For example, the model can simulate the effects of decreasing the viral polymerase rate, which is analogous to adding a non-nucleoside analog polymerase inhibitor or introducing a mutation in 3D^{pol} (20, 21). To model polymerase inhibition, four simulations were run with iterative threefold decreases in the viral polymerase rate (180 – 6.67 kb/h). As expected, decreasing the polymerase rate did not affect the rate of viral entry, indicated by the level of endosomal +ssRNA (Figure 3.8A). However, there was a threshold effect for the +ssRNA and viral protein, with substantial decreases observed at the lowest polymerase rates (Figure 3.8B and 3.8C). The decreased protein abundances arise from less +ssRNA template available for translation, rather than a decrease in the translation rate itself. Similarly, the number of live virions was dramatically decreased at the lower polymerase rates (Figure 3.8D).

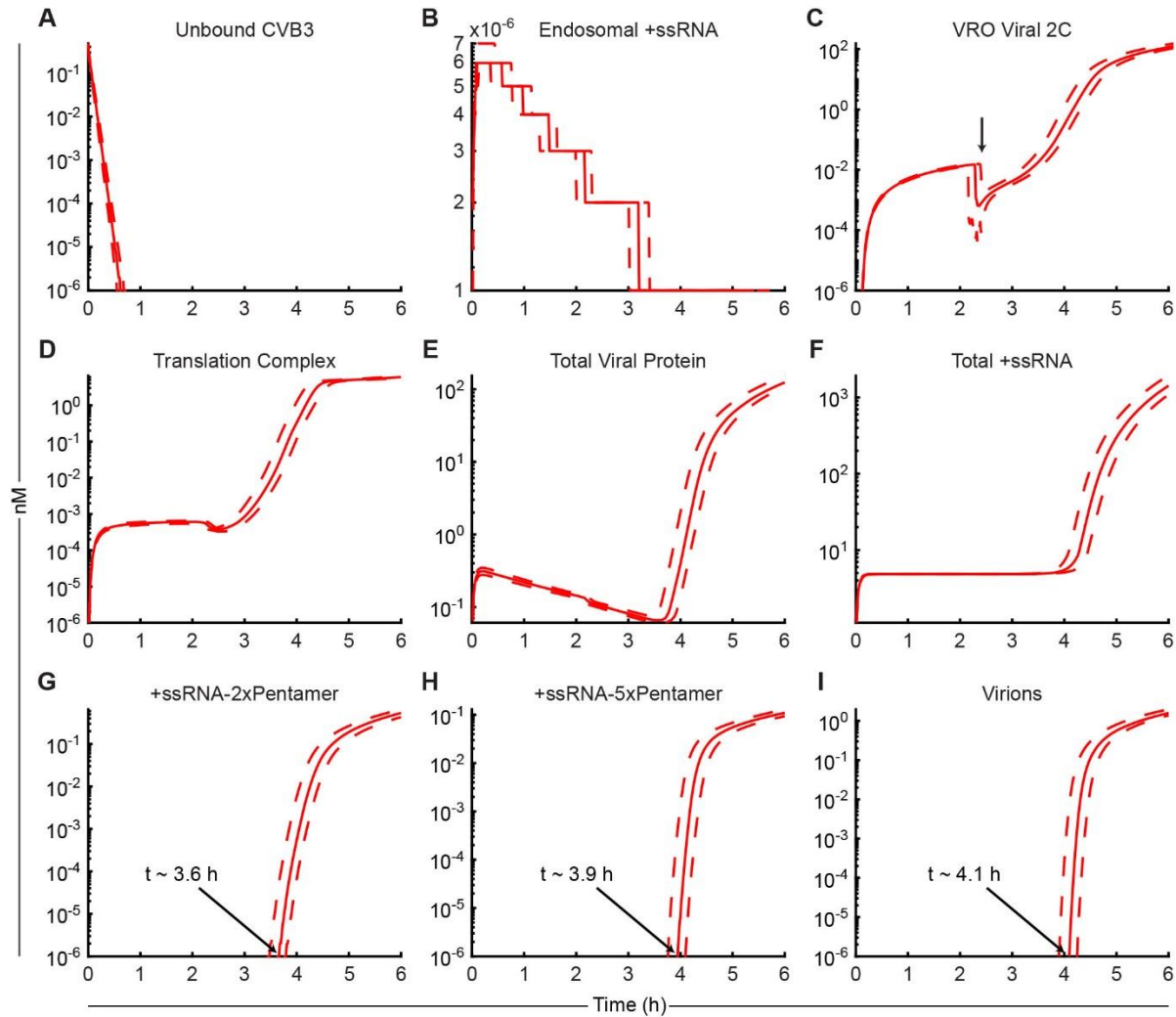


Figure 3.7. Model GUI outputs for the selected species during a viral life cycle.

The simulation was performed for 6 hours with all other parameters set to their default values. (A and B) Species from the delivery module decrease early in the viral life cycle. The concentration of Endosomal +ssRNA (B) is low due to the transient nature of the species. (C–F) Species from the replication module increase over the course of viral infection. Once VROs are formed (C, arrow), the number of translation complexes rapidly increases (D), which causes an explosive increase in viral proteins (E) that can then replicate the viral genome (F). (G–I) Species from the encapsidation module appear in a time-dependent manner. Early RNA–pentamer species (G) appear before intermediate RNA–pentamer species (H), which appear earlier than filled virions (I).

+ssRNA: positive-sense single-stranded viral RNA. VRO: viral replication organelle. +ssRNA- n xPentamer: one +ssRNA bound to n associated pentamers. The simulations were run in the CVB3 model GUI, and the plots were collected using the Export Plot button then processed in Adobe Illustrator.

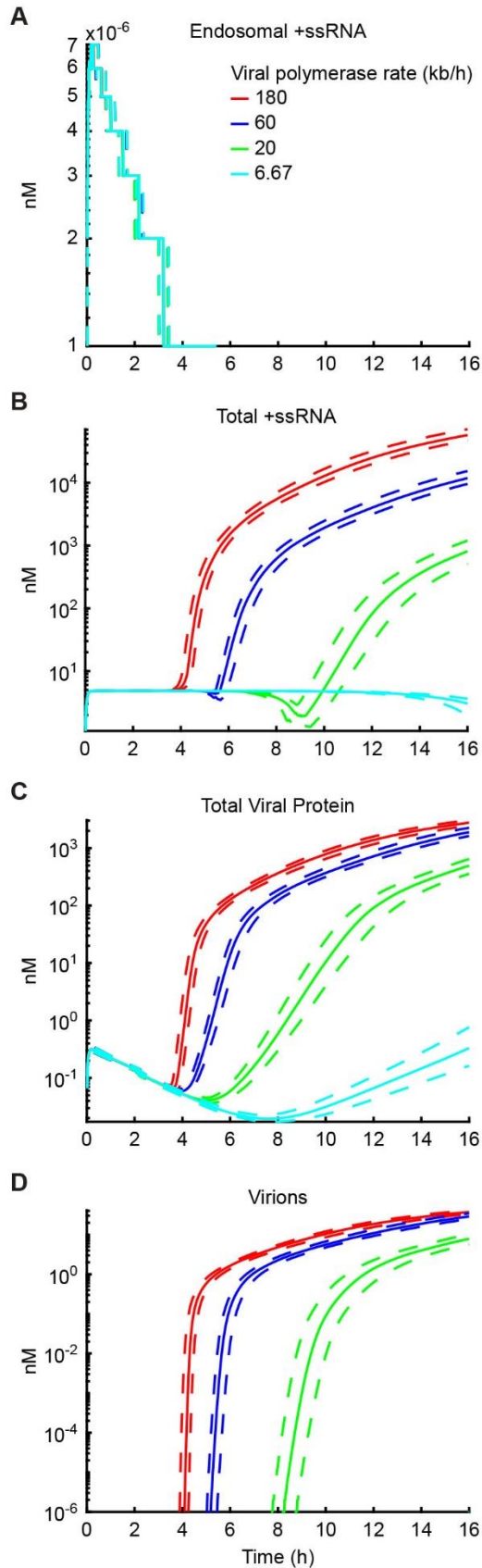


Figure 3.8. Model GUI outputs for selected species identify a threshold viral polymerase rate for productive infection.

Four simulations were performed with iterative threefold decreases in the viral polymerase rate (180 – 6.67 kb/h). All other simulation parameters were set to default. **(A)** Viral entry—measured by endosomal positive-sense single-strand viral RNA (+ssRNA)—is unaffected by the viral polymerase rate. The concentration of Endosomal +ssRNA is low due to the transient nature of the species. **(B–C)** Species in the replication module are sensitive to the viral polymerase rate. The decrease in Total +ssRNA **(B)** decreases the amount of available template for translation of viral proteins **(C)**. **(D)** A threshold viral polymerase rate for productive infection. No virion production is observed for the lowest viral polymerase rate (6.67 kb/h), presumably due to the lack of +ssRNA **(B)** and capsid protein **(C)**. The simulations were run in the CVB3 model GUI, and the plots were collected using the Export Plot button then processed in Adobe Illustrator.

In some instances, it may be more appropriate to perform simulations in population mode rather than single cell mode. One such scenario is at a low MOI. See the **Population mode** section for additional information on the differences in the MOI between single cell mode and population mode. To demonstrate, a simulation was run in both single cell mode and population mode at an MOI of 2. In single cell mode, two virions are not sufficient to cause a productive infection for the given time frame (Figure 3.9A). In population mode, the median agrees with the single cell median, as expected, because the population median represents a single run at an MOI approximately equal to the MOI used in single cell mode (Figure 3.9B). However, some runs in population mode simulate infection for more than two virions since the number of virions in each run is pulled from a Poisson distribution. Therefore, some runs yield a productive infection that is quantifiable when looking at the population-wide average (Figure 3.9B). When using the CVB3 model to conduct experiments, it can be worthwhile to compare single cell mode and population mode to determine if run-to-run variation in MOI affects the experimental outcome.

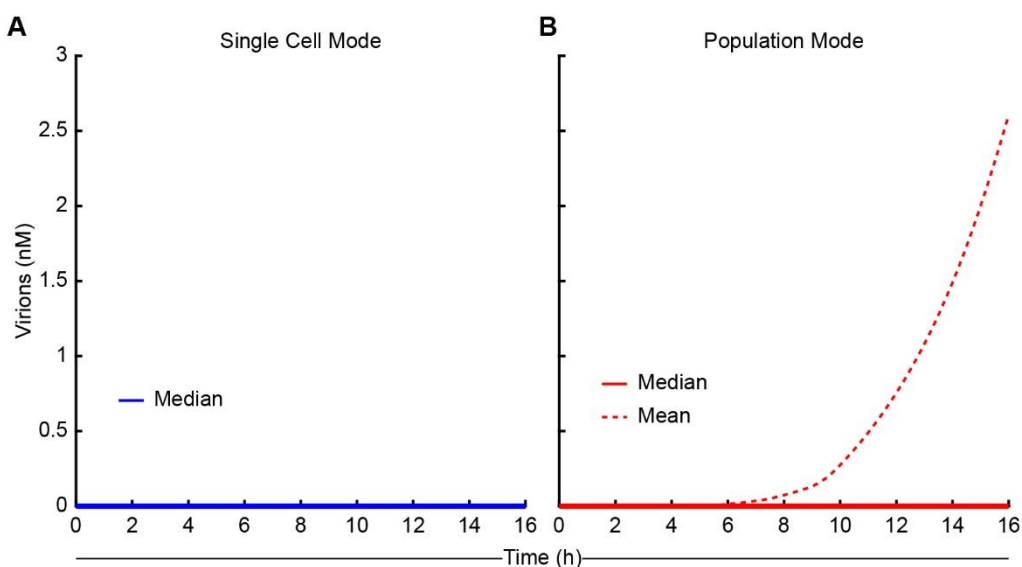


Figure 3.9. Model GUI outputs comparing single cell mode to population mode at a low MOI.

The simulation was performed for 16 hours with MOI set at 2 in either Single Cell mode (A) or Population mode (B) with all other parameters set to their default values. (A) At a low MOI, a single cell does not yield a productive infection (solid blue). (B) In a population of cells, the median response is the same as in a single cell (solid red), while the mean virion production of the entire population yields a quantifiable response (dotted red). The simulations were run in the CVB3 model GUI, and the plots were collected using the Export Plot button then processed in Adobe Illustrator.

3.6 Quantification and statistical analysis

The CVB3 complete kinetics computational model performs repeated runs for a simulation (default: Number of runs = 100). For each run in a simulation, parameter values are drawn from a log-normal distribution. Each log-normal distribution is parameter-specific and centered about the user-defined value of the parameter. The default values are derived from the literature or experiments. The spread of each log-normal distribution is determined by a user-defined coefficient of variation (default: 5%). Drawing from log-normal distributions causes run-to-run variability in output species. Therefore, the data are summarized as the median value \pm confidence interval (default: 90%; determined by the 5th and 95th percentiles of the runs for a simulation). As a result, changing the run count, coefficient of variation, or quantile values will all change the output graphs. For instance, increasing the coefficient of variation will increase the spread of log-normally drawn parameter values and increase the range of output species. In the model GUI, the coefficient of variation is limited to a maximum of 20%, as the model becomes unstable at higher values. Due to differences in the pseudorandom number generation seed, performing the same simulation twice may yield slightly different results. We have found that there is moderate variability between simulations with a per-simulation run count of 10 (Figure 3.10A). By contrast, a run count of 100 is sufficient to limit simulation-to-simulation variability (Figure 3.10B). However, some simulation conditions may be intrinsically more variable and require an increased run count to obtain stable predictions.

Understanding how the model generates uncertainty is also important to draw parallels between model outputs and biology. In the model, it is assumed that every parameter varies a relatively small amount between runs, with all parameters varying to a similar degree. However, in biology these assumptions are not necessarily true, where some features between cells may vary to a great degree (such as protein abundances), while others presumably change very little (such as protein-protein binding rates). Population mode accounts for biological uncertainty in how many live virions infect a cell, but it cannot account for all cell-to-cell variability. As a result, the uncertainty generated by the model should be used to evaluate the variation in model runs and not be interpreted as biological variation.

At the end of the replication cycle, enterovirus virions are released through host-cell bursting. Cell bursting is not encoded in the command-line version of the model, so viral species increase to unrealistic concentrations in some simulation conditions (e.g., over long periods of time). To increase user-friendliness in the GUI, we hardcoded a threshold for the maximum virion concentration based on estimates of the maximum plaque-forming units (PFU) produced by poliovirus in a cell of similar volume to the “cell volume” parameter

in the GUI (22). We converted from PFU to the number of particles using the reported particle-to-PFU ratio (2). Then, we converted the number of particles to a concentration using Equation 1:

$$\text{concentration (nM)} = \frac{x}{N_A} \times \frac{1}{V_{cyto}} \times 10^{24} \quad \text{Equation 1}$$

where x = the number of particles; N_A = Avogadro's number, $6.022 \times 10^{23} \text{ mol}^{-1}$; V_{cyto} = cytoplasm volume in μm^3 , found by subtracting the nuclear volume from the cell volume; and 10^{24} is the conversion factor to obtain concentration in nanomolar (nM).

The calculated maximum virion concentration for the default cytoplasmic volume in the GUI is 326 nM. Beyond the threshold, cell lysis is assumed, and the output data is truncated. Further, we calculated the virion concentration at which cell lysis would be likely to occur based on the median PFU produced by poliovirus (22). The median concentration is 88 nM. Simulated virion outputs greater than 88 nM trigger a warning message that reminds the user that concentrations beyond 88 nM are likely to cause cell lysis.

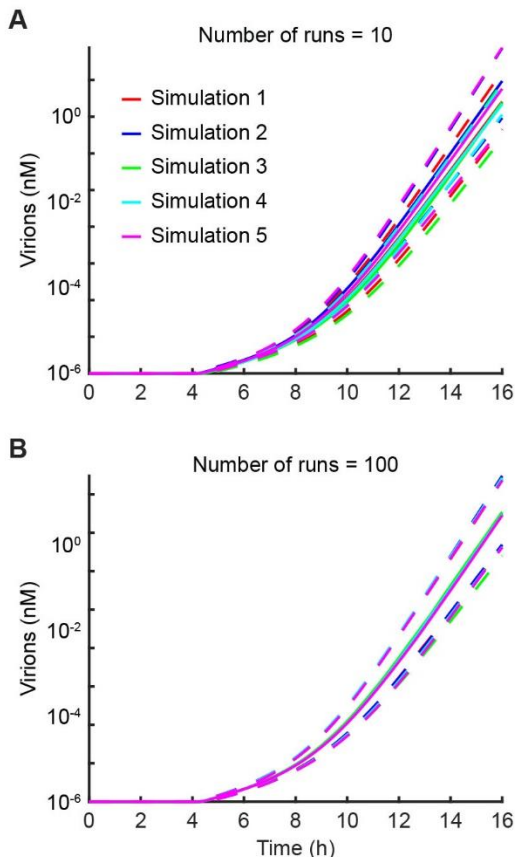


Figure 3.10. Model GUI outputs comparing serial simulations with different run counts.

All other parameters were set to their default values. (A) Five simulations with Number of runs set to 10. (B) Five simulations with Number of runs set to 100. The simulations were run in the CVB3 model GUI, and the plots were collected using the Export Plot button then processed in Adobe Illustrator.

3.7 Limitations

All computational models require assumptions and simplifications, and the CVB3 complete kinetics model is no exception. Assumptions are a necessary trade-off between model accuracy and functionality. Overly complex models often have reduced functionality—increased computational expense and difficult interpretability—and obtain only a marginal increase in molecular detail. Simpler models can be more useful.

One frequently used simplification is lumping a group of reactions or species into a single parameter. For the CVB3 model, three portions of the life cycle were lumped. First, several CVB3 proteins were lumped together. CVB3 encodes two different proteases (2A and 3C), which each have different host cell substrates. In the complete kinetics model, 2A^{pro} and 3C^{pro} are lumped into a single species. Similarly, 3D^{pol} was grouped with the 3AB protein since they strongly interact (23), and three capsid subunits (VP0, VP1, and VP3) were grouped into a single capsid protomer. The remaining hydrophobic proteins were lumped into a single protein (2C^{ATPase}). Second, capsid formation was greatly simplified. Biologically, the binding affinity for subsequent pentamers increases as more capsid pentamers associate and more contact points are available for the next pentamer. Rather than model every possible capsid-forming configuration and its associated binding affinity, the CVB3 computational model neglects geometry and uses the median affinity for every step of capsid formation. Third, the innate immune response is collapsed into a single equation. The full cellular antiviral response to CVB3 consists of a sensor (MDA5) transducing a signal through MAVS and Interferon Regulatory Factor 3 (IRF3) to induce the expression of IFN. IFN is secreted and stimulates the infected cell in an autocrine fashion through the IFN α/β receptor and JAK/STAT pathway to induce the expression of ISGs (15). In the CVB3 model, the entire innate immunity pathway is simplified into a Hill equation, which is fit to ISG expression data from a cardiomyocyte cell line. Despite the simplifications, the complete kinetics model quantitatively captures crucial pieces of the CVB3 life cycle in molecular detail.

The CVB3 complete kinetics model uses mass action to simulate the concentrations of different species in the cell over time. As a result, all output values for measurable species are reported as nanomolar concentrations. Interpreting concentration is difficult for some species, such as virions, which are typically measured in plaque-forming units per milliliter (PFU/mL). The number of PFUs in a viral stock is not reflective of the total number of virions in the stock. In a viral stock, some virions may lack RNA (empty), and others may be incapable of forming plaques (defective). The model generates empty virions, but it does not capture the generation of defective virions. Furthermore, host-cell variations determine

whether a live virion will form a plaque in culture (24). Therefore, model outputs and experimental results involving virions are separate from one another: 1) The complete kinetics model does not account for defective particles, meaning it overestimates the concentration of virions capable of forming a plaque; and 2) Experiments yielding data in PFU/mL do not give any information on the total number of assembled virions. Furthermore, the output virion concentration from the model is the intracellular concentration, since there is no cell lysis component encoded. Instead of a lysis component, a maximum intracellular virion concentration is set in the model, and data beyond the maximum is truncated. Taken together, virion outputs from the model should be treated in relative terms, where the relative change between treatments is more informative than the absolute numbers.

Parameter values in the CVB3 model are set to literature- or experimentally-derived estimates. However, some parameter values are unavailable in the CVB3 literature but are available for closely related viruses; therefore, some default values were derived from literature on other enteroviruses, such as poliovirus. Using parameters derived from other viruses introduces a degree of uncertainty to the model. However, the viral proteins encoded in the model are shared among enteroviruses, and their viral replication cycles are highly conserved. Thus, the parameters likely deviate very little from the actual CVB3 values, lending confidence to the real-world utility of the model.

The CVB3 complete kinetics model GUI functions identically to the command line-based version of the model and contains many adjustable parameters. However, the GUI has some limitations compared to the command line-based CVB3 model. Specifically, 26 parameters are editable using the command line but not in the GUI. These parameters and the rationale for not including them in the GUI are listed in Table 3.4. Additionally, the underlying equations behind the model are rigidly set in the GUI version but can be customized in the command line-based model. This limitation balances user-friendliness and customizability. The GUI version of the model was created to enable any user to run simulations without needing MATLAB installed or knowledge of how to code in MATLAB. If users require the greater range of functions associated with the command line-based model, it is available for download (<https://github.com/JanesLab/CompleteKinetics-CVB3>).

3.8 Troubleshooting

Problem 1:

The simulation program will not open.

Potential solution:

The simulation program will not open if MATLAB Runtime (or the full version MATLAB) is not installed in the correct location. Double check that version 9.10 (R2021a) of MATLAB Runtime is installed. On MacOS systems, MATLAB Runtime must be in the Applications directory in a folder named “MATLAB” (**Install the graphical user interface (GUI), steps 3e and 4a – Figure 3.1**). On Windows systems, MATLAB Runtime must be correctly installed and present in the Add or remove programs list (**Install the graphical user interface (GUI), step 2a**). If MATLAB Runtime cannot be located, download and install the latest version for the appropriate operating system (<https://www.mathworks.com/products/compiler/mcr/index.html>). Alternatively, if the simulation will not open after completing the automatic installation instructions, it may help to instead follow the manual installation instructions.

Problem 2:

The simulation progress bar freezes.

Potential solution:

The simulation will freeze under several circumstances:

1. If the Time (h) box is set too high, the amount of data stored may cause an out of memory error. To avoid this error, reduce the Time (h) value. In test simulations, a Time (h) value of over 50,000 led to model freezing. (The maximum value depends on the specifications of the computer used to run the simulation.) Similarly, exporting data and clearing all active simulations may help if the GUI is running slowly due to a lack of memory (**Basic operations, step 5**).
2. The underlying ordinary differential equation solver encounters model parameters that are difficult to simulate. Avoid exploring input parameter combinations that lead

to model freezing. For example, increasing the +ssRNA degradation rate to 5 and the Time (h) value to 96 can lead to model freezing (**Basic operations, step 6**).

3. Incompatible input parameter values are entered into the model. Input parameters have been constrained in the GUI to avoid incompatibilities with the model. These incompatibilities are the reason why the coefficient of variation has been constrained to a maximum of 20%. However, if simulations are run with the coefficient of variation at or near 20%, freezing due to parameter incompatibility remains a possibility (**Quantification and statistical analysis**). Reduce the coefficient of variation to avoid this problem.
4. Unfortunately, the only solution when the model GUI is completely frozen is to close the GUI and re-open it. All data from previously active simulations will be lost. To prevent this error from occurring in the future, avoid circumstances that lead to conditions prone to freezing.

Problem 3:

No data appear in main GUI plot after running a simulation.

Potential solution:

Check that the Simulation Ready lamp has turned from red to green. The green light indicates that the program has completely finished running and plotting the simulation. If there are still no data in the main GUI plot, uncheck the Log Scale Y checkbox (**Basic operations, step 3**). It is possible that the output species from the simulation is at a concentration of zero throughout the simulation and therefore cannot be plotted when the Y axis is set at a log scale. Using a linear scale, the data will appear as a line along the base of the plot if the species concentration is at zero (Figure 3.3B, green). Alternatively, press the Reset all Parameters to Default Values button (in the Advanced Options tab) and run a default simulation to restore data plotted in the main GUI plot.

Problem 4:

The model displays the same output between two simulations when a parameter is changed, even though different output is expected.

Potential solution:

Depending on the simulation, one of two things may be happening:

1. If the simulation involves an exogenous interferon stimulation (i.e., IFN Stimulation Time field is on), ensure that the Antiviral Response switch is toggled on (**Basic operations, step 4**). If the Antiviral Response switch is off, IFN stimulation will not affect the model simulation. If toggling the Antiviral Response switch on still does not achieve the expected result, see the next potential solution.
2. It is possible that the current set of parameter values overshadows the effect of an additional change. Click the Reset all Parameters to Default Values button (in the Advanced Options tab) to restore the default values and run a baseline simulation to ensure the model is functioning (**Basic operations, step 6**). Then, alter the parameter(s)-of-interest, and perform another simulation.

If the expected output is still not achieved, the chosen parameter values may represent a scenario in which there is truly no change to the system or the differences are so small that the tolerances in the model do not reflect any change.

Problem 5:

The traces in the graphical output are erratic or irregularly shaped.

Potential solution:

Graphical traces will be irregularly shaped when the ordinary differential equation solver encounters very small parameter values that cause discontinuities in the solution. Irregularities are observed under two conditions:

1. Irregularities occur when plotting output species at extremely low concentrations (e.g. Endosomal +ssRNA, **Expected outcomes – Figure 3.7B**). It should be noted that low concentration-based discontinuities are artifactual and should not be interpreted as discrete jumps in species abundance.
2. Discrete changes in species concentrations, such as during the VRO transition, cause artifacts in the traces (**Expected outcomes – Figure 3.7C**). This problem can be mitigated by exercising caution when interpreting graphs with irregular traces. Alternatively, changing altered parameters closer to their default values can improve erratic model outputs.

3.9 Tables

Table 3.1: General Tab Input Options

Input Name	Default Value	Notes
MOI	10	Multiplicity of infection.
Time (h)	16	Timepoint (in hours) to end the simulation.
Antiviral Response	On	Enables/disables the interferon (IFN) response.
Viral Antagonism	On	Enables/disables viral antagonism of the host IFN response.
IFN Stimulation	Off	Enables/disables exogenous IFN stimulation.
Stimulation Time	0	Time of exogenous IFN stimulation. A negative value indicates pre-treatment with IFN.
Results to Plot	Virions	Changes the parameter plotted on the GUI plot. All active simulations will automatically be plotted by changing this.
Other Output		Other output options for Results to Plot.
Start	N/A	Performs the simulation.
Update Plot	N/A	Refreshes the GUI plot. Updates plot with new Other Output.
Log Scale Y	1 (on)	Sets the GUI plot Y axis to logarithmic scale.
Show Quantile Lines	1 (on)	Shows/hides upper and lower quantile lines.
Show All Plots	N/A	Opens 7 figures per active simulation with plots that show all output species.
Clear Data/Plot	N/A	Deletes all active simulations and clears the GUI plot.
Export Plot	N/A	Saves the GUI plot to file.
Export Results	N/A	For each active simulation, the following tables are saved to file: Parameter Output, Median Results, Upper Quantile, Lower Quantile, Mean Results, and Sensitivity Analysis (if applicable).

Table 3.2. Options Tab Input Options

Input Name	Default Value	Notes
ISG concentration for RNA degradation	5	EC50_RNAdeg—the half-effective concentration in nM of interferon-stimulated genes for viral RNA degradation.
Viral protease concentration for viral detector degradation	0.001	EC50_DetectorDeg—the half-effective concentration in nM of viral protease for viral detector degradation.
ISG concentration for viral protease degradation	20	EC50_Protease—the half-effective concentration in nM of interferon-stimulated genes for viral protease degradation.
ISG concentration for inhibition of viral translation	10	EC50_Translate—the half-effective concentration in nM of interferon-stimulated genes for viral translation inhibition.
Simulation Mode	Single Cell	Toggles between a Poisson distribution (population) or using a single integer (single cell) for MOI. Quantile lines are disabled in population mode. Toggling this switch deletes all active simulations and clears the GUI plot.
Sensitivity Analysis	0 (off)	Indicates if a sensitivity analysis should be simulated. During a sensitivity analysis the run count is set to one and quantile lines are disabled. Toggling this switch deletes all active simulations and clears the GUI plot.
Output species	Plus RNA	Changes Sensitivity Analysis output.
Other output		Other output options for Sensitivity Analysis. Multiple species may be entered (separated by commas, with no spaces) and the output will be the sum of the entered species.
Scaling factor	10	Factor to scale parameters by during sensitivity analysis.
Max orders of magnitude	2	Orders of magnitude to scale (in both

to scale		directions) during sensitivity analysis.
Number of runs	100	Indicates the number of runs in a simulation. Runs are not identical, as parameters are stochastically sampled from independent log-normal distributions.
Upper quantile	0.95	The upper quantile of run values to be plotted for indicated output species.
Lower quantile	0.05	The lower quantile of run values to be plotted for indicated output species.
Coefficient of variation	0.05	The coefficient of variation (CV) of the log-normal distribution from which the parameters are stochastically determined. A larger value indicates greater parameter variability between runs. Higher CV values can cause model instability and increase the likelihood of errors. When set at the CV maximum (0.2), approximately one error will occur every 700–800 runs.

Table 3.3. Advanced Options Tab Input Options

Input Name	Default Value	Notes
DAF	4.4e+04	Initial number of unbound cell surface DAF molecules.
CAR	5.5e+06	Initial number of unbound cell surface CAR molecules.
Ribosomes	1e+05	Initial number of host cell ribosomes.
Cell volume	3700	Total volume of the cell in μm^3 . This must be greater than nucleus volume.
Nucleus volume	960	Total volume of the cell nucleus in μm^3 .
Cell surface area	2200	Total outer cell surface area in μm^2 .
VRO surface area	120	Surface area of the average viral replication organelle (VRO) in μm^2 .
CVB3 translocation to tight junction	60	k_transloc—the rate of DAF translocation into and out of the tight junction in h^{-1} . This rate determines

		the speed at which CVB3 is delivered by DAF to bind CAR in the tight junction. The default is set at maximum delivery to CAR in $\alpha 15$ min.
CVB3 internalization	0.5	k_internal—the rate of internalization of the CVB3 bound to CAR in h^{-1} . This also results in downregulation of CAR at the cell surface. The default is set to achieve maximum internalization within ~ 20 min.
Endosome escape	420	k_endo_escape—the rate of CVB3 +ssRNA release from internalized endosomes into the cytoplasm in h^{-1} . Default is set at $\alpha 70$ particles released per 10 min.
Viral polymerase rate	180	Rate of transcription by the viral polymerase for production of + and –ssRNA, in kilobase pairs per hour (kb/h).
Host polymerase rate for ISGs	170	k_transcribeISG—the rate of transcription for host ISGs after viral detection or interferon stimulation, in kilobase pairs per hour (kb/h).
3Dpol molecules required to form VRO	25	Number of 3D ^{pol} molecules required to form a membrane-bound VRO and accelerate viral replication.
Replication complex formation rate: +RNA	0.36	k_R_Ip_Form—the rate of viral +ssRNA binding to 3D ^{pol} in a VRO to form a replication complex in $nM^{-1} h^{-1}$.
Replication complex formation rate: -RNA	0.36	k_R_In_Form—the rate of viral –ssRNA binding to 3D ^{pol} in a VRO to form a replication complex in $nM^{-1} h^{-1}$.
Translation rate	3.5	The translation rate of CVB3 +ssRNA into viral polyprotein, in codon/s.
Host ribosome cleavage rate	1158	kcat_cleave—the rate of host ribosome (EIF4G) cleavage by viral proteases in h^{-1} . The cleavage reaction shifts host ribosomes to viral ribosomes.
Protease Km for host ribosomes	960e+03	Km_cleave—the Km value, indicating the affinity of viral protease for host ribosome (EIF4G) in nM. The cleavage reaction shifts host ribosomes to viral

		ribosomes.
Translation complex formation rate	25/Polysome size	$k_{T_c_Form}$ —the rate of viral +ssRNA binding to ribosomes to form translation complexes in $nM^{-1} h^{-1}$. Polysome size is set at 2.5.
Pentamer assembly rate	3.6	The capsid:capsid (with or without viral RNA) and capsid:RNA rate of association in $nM^{-1} h^{-1}$. This parameter simultaneously changes k_{1f_cap} , k_{2f_cap} , and $k_{RNA_CapBind}$.
Pentamer disassembly rate	3600	The capsid:capsid (with or without viral RNA) and capsid:RNA rate of dissociation in h^{-1} . This parameter simultaneously changes k_{1b_cap} , k_{2b_cap} , and $k_{RNA_CapUnbind}$.
+ssRNA degradation rate	0.23	The basal rate of positive-strand viral RNA degradation in h^{-1} . This parameter simultaneously changes u_{P_cyt} and u_{P_VRO} .
-ssRNA degradation rate	0.23	The basal rate of negative-strand viral RNA degradation in h^{-1} . This parameter simultaneously changes u_{N_cyt} and u_{N_VRO} .
Translation complex degradation rate	0	u_{T_c} —the degradation rate for +ssRNA in translation complexes in h^{-1} . The default is set to zero under the assumption that Poly(rC) binding proteins protect the viral RNA during polysome formation.
Viral protein degradation rate	0.05	The basal rate of viral 3D ^{pol} , protease, and capsid degradation in h^{-1} . This parameter simultaneously changes $u_{VirProt_cyt}$, $u_{VirProt_VRO}$, u_{cap_cyt} , and u_{cap_VRO} .
ISG degradation rate	0.1	u_{ISG} —the rate of ISG protein degradation by the host cell in h^{-1} .

Table 3.4. Parameters from the complete kinetic model that are excluded from the CVB3 model GUI.

Name	Description	Notes
RibAvail	Initial concentration of ribosomes	This parameter is dependent

	available to the virus and not host.	on the total number of ribosomes.
RibUnavail	Initial concentration of ribosomes available to the host and not virus.	This parameter is dependent on the total number of ribosomes.
CVB3diameter	Diameter of a CVB3 virion.	
Pol3DLength	Length of the CVB3 3D polymerase.	
ISGbpLength	Length of ISG15 in kilobase pairs, which is used to determine the rate of ISG formation.	Changing the host polymerase rate for ISGs would have a similar effect as changing this parameter.
RiboActive	Proportion of all ribosomes that are currently active in the cell.	Changing the total number of ribosomes would have a similar effect as changing this parameter.
PolysomeSize	Average number of ribosomes bound to a single mRNA during translation.	
CVB3genomelength	Length of the CVB3 genome in nucleotides.	
CVB3polyprolength	Length of the CVB3 polyprotein in amino acids.	
VROVolConv	Volume conversion to model the concentrating effect of VROs.	Parameter is dependent on cell volume, nucleus volume, and VRO surface area.
k_on_DAF	Association rate between CVB3 and DAF.	Changing the DAF concentration would have a similar effect as changing this parameter.
k_off_DAF	Dissociation rate between CVB3 and DAF.	
k_on_CAR	Association rate between CVB3 and CAR.	Changing the CAR concentration would have a similar effect as changing

		this parameter.
k_off_CAR	Dissociation rate between CVB3 and CAR.	
k_Translate	Overall rate of CVB3 translation.	Calculated from the translation rate, polysome size, and polyprotein length.
k_N_Transcript	Rate of CVB3 negative-strand synthesis.	Calculated from the viral polymerase rate and CVB3 genome length.
k_P_Transcript	Rate of CVB3 positive-strand synthesis.	Calculated from the viral polymerase rate and CVB3 genome length.
k_P_on	Rate of CVB3 positive-strand translocation to the VRO.	
k_P_off	Rate of CVB3 positive-strand translocation out of the VRO.	
k_N_on	Rate of CVB3 negative-strand translocation to the VRO.	
k_N_off	Rate of CVB3 negative-strand translocation out of the VRO.	
k_Pentamer_on	Rate of CVB3 capsid pentamer translocation to the VRO.	
k_Pentamer_off	Rate of CVB3 capsid pentamer translocation out of the VRO.	
n_Hill	Hill coefficient used to model viral detection by the host cell.	Fit to ISG expression data from cardiomyocytes.
kD_Hill	Dissociation constant used in the Hill equation to model viral detection by the host cell.	Fit to ISG expression data from cardiomyocytes.
ISGForm	Formation rate of ISG proteins.	Changing the host polymerase rate for ISGs would have a similar effect as changing this parameter.

3.10 References

1. F. J. Bruggeman, H. V. Westerhoff, The nature of systems biology. *Trends Microbiol* **15**, 45–50 (2007).
2. A. B. Lopacinski, *et al.*, Modeling the complete kinetics of coxsackievirus B3 reveals human determinants of host-cell feedback. *Cell Syst* **12**, 304-323.e13 (2021).
3. V. Racaniello, “Picornaviridae: the viruses and their replication” in *Fields Virology*, 6th ed, D. M. Knipe, P. M. Howley, Eds. (Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013).
4. D. R. Shafren, *et al.*, Coxsackieviruses B1, B3, and B5 use decay accelerating factor as a receptor for cell attachment. *J Virol* **69**, 3873–3877 (1995).
5. J. M. Bergelson, *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320–1323 (1997).
6. C. B. Coyne, J. M. Bergelson, Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. *Cell* **124**, 119–131 (2006).
7. B. J. Kempf, D. J. Barton, Poliovirus 2A(Pro) increases viral mRNA and polysome stability coordinately in time with cleavage of eIF4G. *J Virol* **82**, 5847–5859 (2008).
8. C. E. Melia, *et al.*, Origins of Enterovirus Replication Organelles Established by Whole-Cell Electron Microscopy. *mBio* **10**, e00951-19 (2019).
9. S. Miller, J. Krijnse-Locker, Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**, 363–374 (2008).
10. P. Jiang, Y. Liu, H.-C. Ma, A. V. Paul, E. Wimmer, Picornavirus morphogenesis. *Microbiol Mol Biol Rev* **78**, 418–437 (2014).
11. J. Andrejeva, *et al.*, The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. *Proc Natl Acad Sci U S A* **101**, 17264–17269 (2004).
12. D. Kang, *et al.*, mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A* **99**, 637–642 (2002).
13. T. Kawai, *et al.*, IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol* **6**, 981–988 (2005).
14. R. B. Seth, L. Sun, C.-K. Ea, Z. J. Chen, Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).
15. A. J. Sadler, B. R. G. Williams, Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559–568 (2008).

16. Q. Feng, *et al.*, Enterovirus 2Apro targets MDA5 and MAVS in infected cells. *J Virol* **88**, 3369–3378 (2014).
17. A. Mukherjee, *et al.*, The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog* **7**, e1001311 (2011).
18. D. Etchison, S. C. Milburn, I. Edery, N. Sonenberg, J. W. Hershey, Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 220,000-dalton polypeptide associated with eucaryotic initiation factor 3 and a cap binding protein complex. *J Biol Chem* **257**, 14806–14810 (1982).
19. E. L. Ellis, M. Delbrück, THE GROWTH OF BACTERIOPHAGE. *J Gen Physiol* **22**, 365–384 (1939).
20. L. van der Linden, *et al.*, The RNA template channel of the RNA-dependent RNA polymerase as a target for development of antiviral therapy of multiple genera within a virus family. *PLoS Pathog* **11**, e1004733 (2015).
21. M. Wang, *et al.*, Stringent control of the RNA-dependent RNA polymerase translocation revealed by multiple intermediate structures. *Nat Commun* **11**, 2605 (2020).
22. T. H. Dunnebacke, M. B. Reaume, Correlation of the yield of poliovirus with the size of isolated tissue cultured cells. *Virology* **6**, 8–13 (1958).
23. W. Xiang, A. Cuconati, D. Hope, K. Kirkegaard, E. Wimmer, Complete protein linkage map of poliovirus P3 proteins: interaction of polymerase 3Dpol with VPg and with genetic variants of 3AB. *J Virol* **72**, 6732–6741 (1998).
24. K. C. M. Goh, *et al.*, Molecular determinants of plaque size as an indicator of dengue virus attenuation. *Sci Rep* **6**, 26100 (2016).

Chapter 4:

Proteome-wide copy-number estimation from transcriptomics[#]

Andrew J. Sweatt¹, Cameron D. Griffiths¹, B. Bishal Paudel¹, Kevin A. Janes^{*1,2}

¹Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA, 22908

²Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, USA, 22908

*Corresponding Author: Kevin A. Janes, Box 800759 Health System, University of Virginia, Charlottesville, VA 22908. Phone: 434-243-2126. Email: kjanes@virginia.edu

[#]This work has been pre-printed on *bioRxiv*.

<https://www.biorxiv.org/content/10.1101/2023.07.10.548432v1.full>.

Tables referenced in this text are too large to be included in this dissertation. Please refer to the source publication to obtain them.

4.1 Foreword

My second major project seems, admittedly, unrelated to my first. Thus, I would like to briefly document its history here. When constructing the entry module for the CVB3 model, we lacked the value for DAF abundance. To estimate its abundance, we developed a method that fit a hyperbolic-to-linear equation through paired mRNA–protein abundance measurements for DAF from 14 HeLa derivatives (Figure 2.3D). During revisions of the *Cell Systems* paper describing the CVB3 model, we became aware of a new proteomic dataset from Steve Gygi’s group—the proteomics that ultimately became our training data for this project. The proteomics were collected in cell lines in the Cancer Cell Line Encyclopedia, which also had RNA-seq data. We saw this paired raw data as a great opportunity for determining whether gene-specific models that predict protein abundance from mRNA abundance could be built.

At the same time, I became interested in using the CVB3 model to make predictions about susceptibility to CVB3 infection based on the abundances of DAF and CAR. CAR had previously been shown to be important for infection, but the conditional abundance of CAR on DAF had not been shown. To build individualized models of CVB3 susceptibility, I needed a way to estimate protein copy numbers with better accuracy than we had previously done. With a biological question in mind and raw data at my fingertips, this seemingly unrelated project came into existence, and I am happy to share it with you now.

4.2 Abstract

Protein copy numbers constrain systems-level properties of regulatory networks, but absolute proteomic data remain scarce compared to RNA-seq. We related mRNA to protein statistically using best-available data from quantitative proteomics-transcriptomics for 4366 genes in 369 cell lines. The approach starts with a central estimate of protein copy number and hierarchically appends mRNA-protein and mRNA-mRNA dependencies to define an optimal gene-specific model linking mRNAs to protein. For dozens of cell lines and primary samples, these protein inferences from mRNA outmatch stringent null models, a count-based protein-abundance repository, and empirical protein-to-mRNA ratios. The optimal mRNA-to-protein relationships capture biological processes along with hundreds of known protein-protein complexes, suggesting mechanistic relationships. We use the method to estimate viral-receptor abundances from human heart transcriptomes and build 1489 systems-biology models of coxsackievirus B3 infection susceptibility. When applied to 796 RNA-seq profiles of breast cancer, inferred copy-number estimates collectively reclassify 26-29% of luminal tumors. By adopting a gene-centered perspective of mRNA-protein covariation across different biological contexts, we achieve accuracies comparable to the technical reproducibility of contemporary proteomics.

4.3 Introduction

Absolute numbers of molecules place important bounds on biological systems, but they are hard to come by (1). One exception is deep RNA sequencing (RNA-seq) of bulk samples, which provides absolute transcript-per-million (TPM) estimates of all expressed genes (2). The commoditization of sequencing has made RNA-seq the prevailing omics approach: as of mid-2023, the leading repository (3) contains >32,000 studies with human samples. RNA-seq profiles are useful for reading out the state of the genome (4, 5), but mapping the transcriptome to the abundance of proteins is complex. In tumor classification, for example, the number and identity of cancer subtypes changes when using quantitative measurements of mRNA versus protein (6, 7). The challenge is especially acute for mathematical models in systems biology, which need protein quantities to constrain topology, initial conditions, or transition rates (8–10). Filling the overall gap requires new strategies for absolute quantification of proteomes for different needs.

Progressive experimental innovations in untargeted mass spectrometry have made quantitative proteomics a reality (11). Isobaric labeling approaches such as tandem mass tagging (TMT) now quantify 11+ multiplex samples and are the method of choice for proteogenomics (12, 13). Comparisons of individual proteins across samples indicate that linear mRNA–protein relationships vary greatly in quality (Pearson $R = -0.4$ to 0.8)

depending on the gene and gene category (6, 7). Unfortunately, multiplex labeling yields peptide-specific relative quantities that cannot examine absolute differences among proteins within a sample (11). This difficulty is surmounted by data-independent acquisition methods like sequential window acquisition of all theoretical mass spectra (SWATH), which analyzes all precursor ions in a series of mass-to-charge ratio windows (14). After data acquisition, the most sensitive peptide(s) of a protein are summed by intensity, and the resulting data are centered at a reasonable per-cell average (10^4 copies) to yield absolute estimates of the detectable proteome. SWATH is newer, harder to adapt to different cell lineages, and lower throughput. Consequently, the leading proteomics repository contains ~tenfold fewer SWATH depositions than label-based depositions as of mid-2023 (15). The need for commoditized SWATH-like protein estimates may forever outpace the ability to generate them directly.

As a means for bootstrapping absolute protein copy numbers, an appealing starting point is RNA-seq. mRNA is the template for protein translation, and in terms of scale, depositions of human RNA-seq exceed those of quantitative proteomics (all species, all methods) by ~fivefold (3, 15). However, despite useful transcriptomic inference of protein activities from the gene networks surrounding them (16), directly estimating absolute protein copy numbers from mRNA is historically fraught with uncertainty. Linear mRNA–protein relationships adequately recapitulate protein expression among genes within a sample, but they are poor at distinguishing protein differences among samples for any given gene (17–20). The latter is important for systems biology when using transcriptome profiles to instantiate personalized models of function (8–10). The current thinking is that the steady-state abundance of mRNA and its intrinsic translation rate create a general “set point” for protein expression, which is buffered or tuned according to the abundance of complexes that stably contain the protein (17, 21). Unfortunately, our working inventory of protein-protein interactions and stable complexes in mammalian cells is far from saturation (22, 23), which has thus far prevented a bottom-up reconstruction of mRNA-to-protein relationships that are absolute and conditional.

Here, we surmounted this challenge by adopting a top-down perspective that statistically identifies the best working absolute mRNA-to-protein relationship for each gene based on paired data in several hundred cancer cell lines. SWATH and TMT datasets from different sources are encouragingly self-consistent, enabling the meta-assembly used for model training and selection of three relationship classes. Although relationship classes are entirely data driven, we find biological meaning and gene-specific mechanisms in each. The approach consistently improves the accuracy of proteome-wide inferences from RNA-seq transcriptomes of cells, tumors, and tissues when compared to other tools

and a stringent null hypothesis specific to each gene's protein set point. We use the method to build 1489 personalized systems-biology models of virus infection (24, 25) and re-classify 796 cases of breast cancer (26) according to inferred absolute protein abundance from public RNA-seq datasets. This study provides an open and accessible route to gleaning protein copy numbers from RNA-seq when it is impractical or impossible to quantify the proteome directly (<http://janeslab.shinyapps.io/Pinferna>).

4.4 Results

4.4.1 Deriving three gene-specific biological classes of mRNA-protein relationships

To estimate mRNA-protein relationships, we obtained quantitative proteomics measured by TMT mass spectrometry in 375 cancer cell lines (27) and placed these data on an absolute scale by using independent SWATH proteomics from two lines—one breast carcinoma (28) and one osteosarcoma (29)—within the TMT dataset (Figure 4.1A, Step 1, Figure 4.2A and Dataset EV1). Training with a large, diverse panel of cancer lines avoids confounding gene or protein covariations that may arise in primary tissues and tumors because of cell mixtures (30). When TMT profiles scaled to one reference line were compared to SWATH data measured directly in the other reference line, correlations were above 0.7 in both cases ($p < 10^{-15}$; Figure 4.1B and Figure 4.2B), placing bounds on the internal consistency of the two data sources. Overall, the meta-assembly yielded protein copy number measurements for 4384 proteins across 375 cell lines.

We discerned mathematical relationships that best captured abundance relationships between mRNA and protein by merging SWATH-scaled proteomics with paired Cancer Cell Line Encyclopedia transcriptomics (31) and building gene-specific regressions of different classes (Figure 4.1A, Step 2 and Datasets EV2–EV4). In the simplest case, protein abundance varies nominally around the median (M) regardless of transcript abundance (Figure 4.1C). To incorporate abundance information about the transcript, we also evaluated a hyperbolic-to-linear (HL) relationship where low-abundance changes in mRNA cause larger nonlinear changes in protein that linearize as mRNA abundance increases (Figure 4.1D). Finally, we considered that the abundance of some proteins may be further explained by the abundance of other transcripts and applied the least absolute shrinkage and selection operator (LASSO) to residuals of the HL fit (Figure 4.1E). The best model among the three for each gene was distinguished by the Bayesian Information Criterion (BIC; Figure 4.1A, step 3 and Figure 4.2C–E) to arrive at preferred M, HL, or HL+LASSO relationships for 4366 genes (Dataset EV5). The best model for each gene was

strongly preferred over the others in 98% of cases (Figure 4.2F). These relationships create a template for protein inference from RNA (Pinferna) given new samples with transcriptomic profiles (Figure 4.1A, step 4).

We examined characteristics of the genes in each model class. Consistent with previous findings (7, 32), M genes with no clear transcript dependence showed gene ontology (GO) enrichments for translation and mitochondrial electron transport (Figure 4.3A and Dataset EV6). Although M genes were not significantly longer lived than others (Figure 4.4A and 4.4B), we found that they had high transcript abundances (Figure 4.4C) and were enriched in multi-protein complexes in the CORUM database ($p < 10^{-59}$) (33). Proteins residing in stable complexes may saturate for all measured abundances of mRNA because their copy numbers are stoichiometrically limited by other subunits (21), causing the loss of an observable relationship between mRNA and protein.

Most genes exhibited some dependence on their mRNA (Figure 4.1D and 4.1E). To incorporate transcript information, we assessed various low-complexity models involving one (linear), two (hyperbolic) or three (3-parameter logistic or HL) free parameters. The four models were compared by BIC, and HL was overwhelmingly the best or near-best model for most genes (Figure 4.3B). Results were similar when using BIC weights (34) to assess the relative likelihood of HL against the others (Figure 4.4D). HL accommodated rare log-concave down relationships that occurred when protein abundance saturated at high transcript abundances ($p < 10^{-15}$; Figure 4.3C). Loss of transcript dependence arises biologically when a protein subunit surpasses the abundance of the complex in which it resides (21, 35), as observed for M genes across their entire measured range of mRNA. Accordingly, among HL genes, those that were log-concave down were mildly enriched in protein complexes in the CORUM database ($p < 0.05$) (33). More common were log-concave up HL relationships in which protein abundance increased only at higher mRNA abundances (Figure 4.3D). Using a simple computational model for synthesis and turnover of mRNA and protein with dimensionless rate-parameter estimates, we found that log-concave up relationships arose naturally when steady-state abundances of mRNA and protein were halved and randomly sampled along the trajectory back to steady state (Figure 4.4E and 4.4F). Such “halving-and-random-sampling” occurs when cells asynchronously undergo cytokinesis, halving the protein copies per cell and sporadically re-entering into G1. Other HL genes showed mixed concavities or were linear to different extents (Figure 4.4G–I). Taken together, HL regressions provided the flexibility needed to capture various biological mechanisms that relate mRNA to protein (Figure 4.1D).

One-third of HL regressions were statistically improved by adding mRNA features selected and weighted using LASSO (Figure 4.1E). LASSO features were enriched for

cytoplasmic translation (GO:0002181; $q < 10^{-10}$) and the proteasomal pathway (GO:0043161; $q < 10^{-10}$), suggesting dependencies that may promote protein synthesis or turnover. Specific *RPS*- or *RPL*-prefixed transcripts of the ribosome and *PSM*-prefixed transcripts of the proteasome were also enriched among LASSO-selected features ($p_{\text{ribosome}} < 10^{-109}$, $p_{\text{proteasome}} < 10^{-18}$). Ribosome feature weights were disproportionately positive, whereas proteasome subunits were negative (Figure 4.3E), consistent with their expected influence on protein abundance. Overall, we asked whether LASSO-selected genes were more likely to interact physically with the protein of interest. Using the STRING database (36), we found a remarkable enrichment for interactions among LASSO-selected genes ($p \ll 10^{-4}$; Figure 4.3F), indicating that features contain more than spurious statistical associations. We conclude that Pinferna's three-tiered modeling approach captures various biological phenomena and mechanisms in its framework.

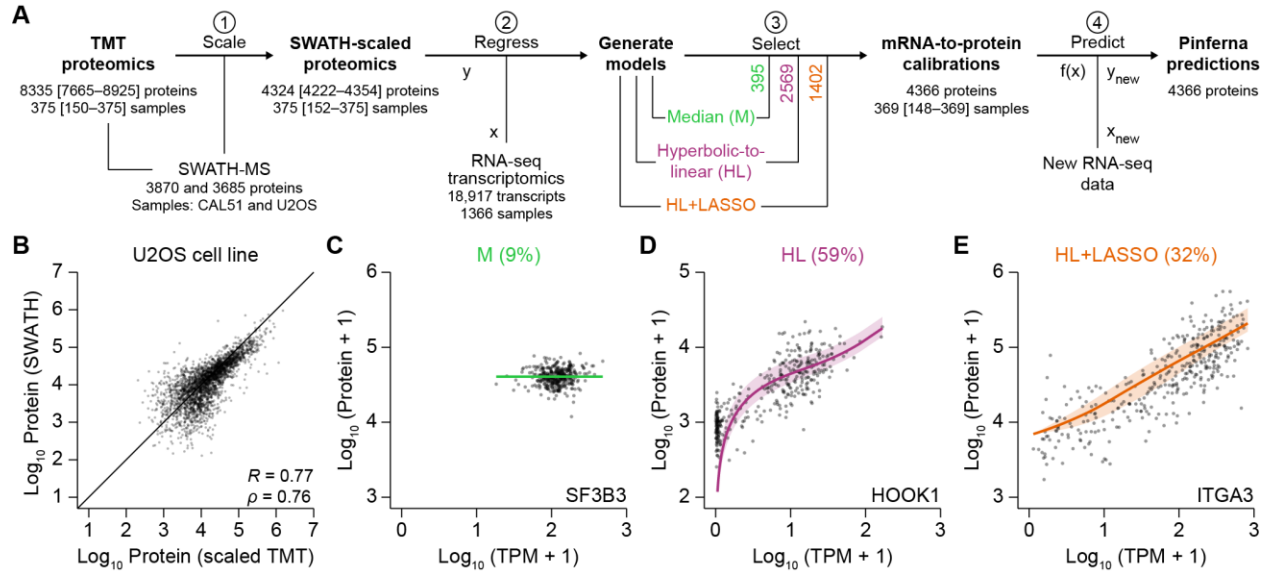


Figure 4.1. Meta-assembly and inference of conditional mRNA-to-protein relationships for 4366 human genes.

(A) Data fusion and model discrimination. (1) Tandem mass tag (TMT) proteomics of 375 cancer cell lines (27) were calibrated to an absolute scale based on sequential window acquisition of all theoretical mass spectra (SWATH) proteomics of CAL51 and U2OS cells (PXD003278; PXD000954). (2) SWATH-scaled proteins were regressed using three models that incorporate transcript abundance from RNA sequencing (RNA-seq) to different extents: median (M), no contribution of mRNA; hyperbolic-to-linear (HL) relationship incorporating mRNA of the gene, $a \cdot \left(\frac{b \cdot \text{mRNA}}{c + \text{mRNA}} + \text{mRNA} \right)$; HL + least absolute shrinkage and selection operator (LASSO) regressors with mRNAs other than the gene of interest. (3) Model selection for each gene was based on the Bayesian Information Criterion. The number of genes selected in each class is indicated. (4) New samples profiled by RNA-seq were used with the calibrated models to make protein inference from RNA (Pinferna) predictions. The number of proteins measured per sample or number of samples with data per protein is shown at each step as the median with the range in brackets.

(B) Reliable cross-calibration of the TMT and SWATH meta-assembly. Step 1 of Figure 4.1A was performed with CAL51 data alone and the SWATH-scaled TMT proteomics of U2OS cells compared with data obtained directly by SWATH. Pearson's R and Spearman's ρ are shown. The reciprocal cross-calibration is shown in Figure 4.2B.

(C–E) Representative M, HL, and HL+LASSO genes. Absolute protein copies per cell were regressed against the mRNA abundance normalized as transcripts per million (TPM). Best-fit calibrations \pm 95% confidence intervals are overlaid on the proteomic–transcriptomic data from $n = 369$ cancer cell lines. Evidence for model selection is shown in Figure 4.2C–E.

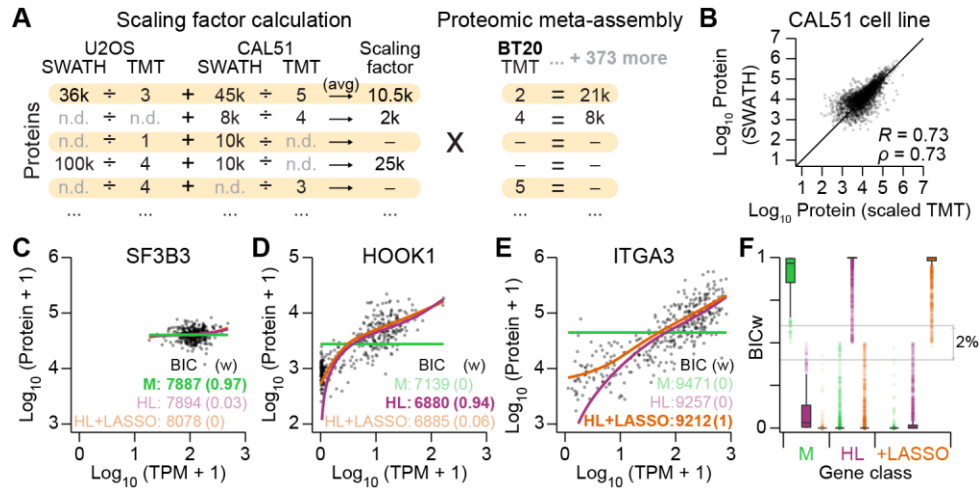


Figure 4.2. Calibration for the proteomic meta-assembly and model selection examples.

(A) SWATH–TMT scaling factor estimation and calibration of the proteomic meta-assembly. Scaling factors were estimated for each protein by dividing the SWATH intensity by the TMT intensity in U2OS and CAL51 when both data types were available and averaging (avg) the two ratios when possible. The gene-specific scaling factors were used to convert the entire TMT dataset (bold) to absolute protein abundances.

(B) The reciprocal cross-calibration to that shown in Figure 4.1B. Step 1 of Figure 4.1A was performed with U2OS data alone and the SWATH-scaled TMT proteomics of CAL51 cells compared with data obtained directly by SWATH. Pearson’s R and Spearman’s ρ are shown.

(C–E) Model selection of the representative genes shown in Figure 4.1C–E. Absolute protein copies per cell were regressed against the mRNA abundance normalized as transcripts per million (TPM) for $n = 369$ cancer cell lines. Data are fit with M, HL, and HL+LASSO models. The BIC was used to discriminate the best model for each fit, and BIC weights (w) are shown in parentheses to indicate the relative best-model probability. The lowest BIC is indicated.

(F) BIC weights (BICw) for each model fit to M, HL, or HL+LASSO (+LASSO) genes. Range of weights for genes with model ambiguity (two models with a BICw ≥ 0.4) are boxed (gray) with the percentage indicated.

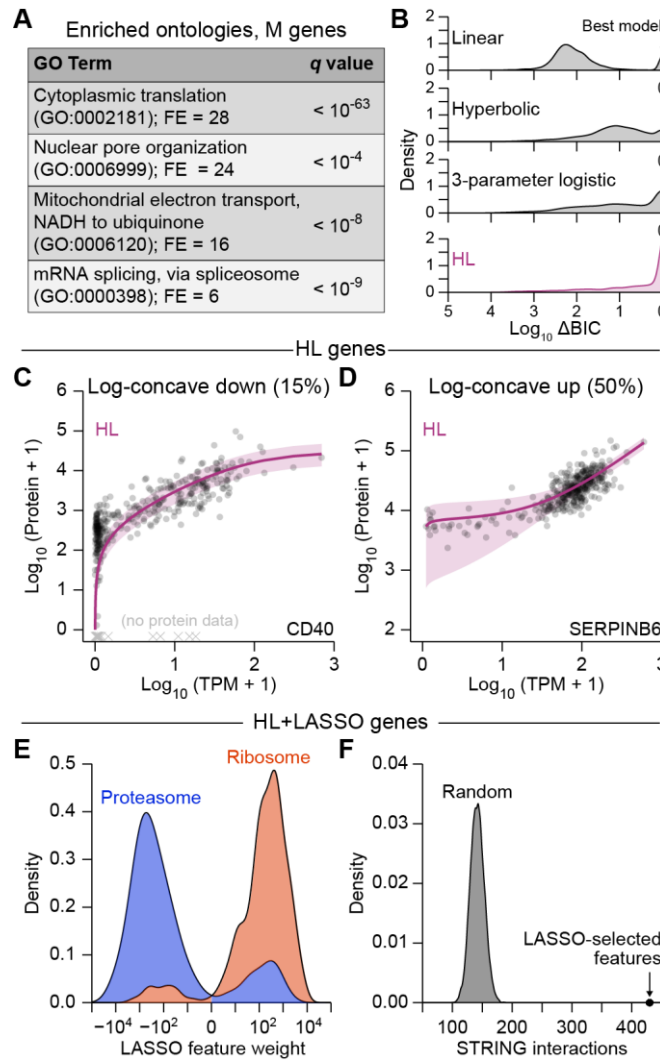


Figure 4.3. Pinferna model selection is consistent with known biological mechanisms and mRNA-to-protein relationships.

(A) Gene ontology (GO) enrichments for M genes. The largest non-redundant GO term is shown with the fold enrichment (FE) and false discovery rate-corrected p value (q). The complete list of GO enrichments for each relationship class is available in Dataset EV6.

(B) HL outperforms competing mRNA-to-protein relationships. Models encoding linear, hyperbolic, three-parameter logistic, and HL relationships were built for all genes ($n = 4366$) and compared by Bayesian Information Criterion (BIC). Results are shown as the smoothed density of BIC differences (ΔBIC) relative to the best model for that gene ($\Delta BIC = 0$). Distributions of BIC weights (34) are shown in Figure 4.4D.

(C, D) HL captures different empirical classes of mRNA-to-protein relationships. Log concave-down genes (C) saturate at high mRNA abundance, whereas log concave-up genes (D) plateau at low mRNA abundance. The remaining genes exhibited characteristics of both fits or linear relationships to varying degrees (Figure 4.4G–I).

(E) Feature weights of HL+LASSO genes are biologically sensible. Smoothed densities of LASSO feature weights (indicating strength and direction of modulation for an HL fit) among mRNAs encoding subunits of the proteasome ($n = 127$ feature weights; blue) and the ribosome ($n = 397$ feature weights; red) are shown.

(F) HL+LASSO feature weights are highly enriched for STRING interactions. For each HL+LASSO gene, LASSO-selected features were replaced with random genes ($n = 10,000$ iterations) to build a null distribution for finding binary interactions in STRING (36). The actual number of STRING interactions among HL+LASSO genes of Pinferna is indicated.

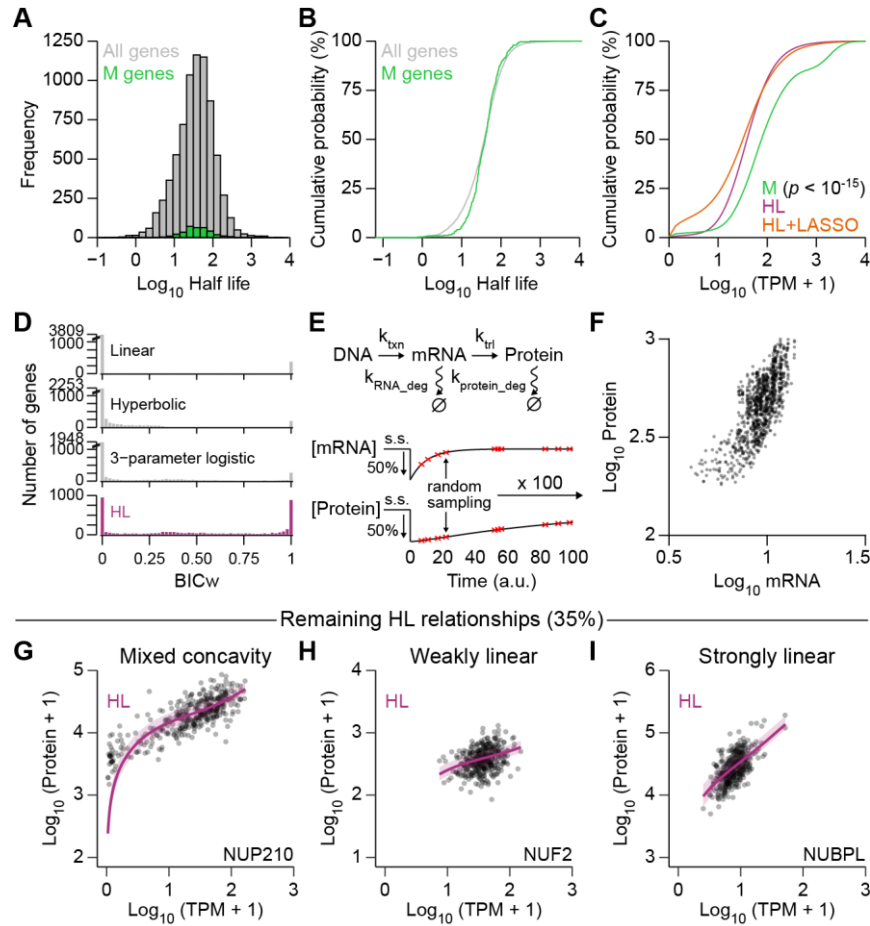


Figure 4.4. Extended characterization of M and HL mRNA-to-protein relationship classes.

(A, B) M genes do not exhibit longer half-lives compared to other relationship classes. Half-lives for proteins were obtained from (37) and plotted by frequency (A) or as an empirical cumulative distribution function (B) for all genes ($n = 7029$ genes (37); gray) and M genes ($n = 334$ genes; green).

(C) M genes are more abundant by mRNA compared to other relationship classes. mRNA abundance was normalized as TPM and placed on a log scale with a pseudocount of 1. Cumulative distribution functions are shown for M genes ($n = 395$ genes; green), HL genes ($n = 2569$ genes; purple), and HL+LASSO genes ($n = 1402$ genes; orange). Distributions were compared by K-S test with Šidák correction for multiple-hypothesis testing.

(D) Distribution of BIC weights (BICw) for models encoding linear, hyperbolic, three-parameter logistic, and HL relationships ($n = 4366$ genes) shown in Figure 4.3B.

(E, F) Log concave-up patterns arise when mRNA and protein abundances recover from transient perturbations to steady-state values. A simple transcription–translation model (E) was reduced by 50% and randomly sampled at 10 time points (red) during the return to steady state (s.s.). For the model, the following dimensionless rate parameters were used: $k_{txn} = 1$; $k_{trl} = 1$; $k_{RNA_deg} = 0.1$; $k_{protein_deg} = 0.01$. The model was simulated 100 times with lognormally distributed parameter noise (coefficient of variation = 10%) and the joint observation of mRNA and protein abundances ($n = 10$ time points \times 100 simulations) is shown in F.

(G–I) Examples of other HL relationships not shown in Figure 4.3C and 4.3D: mixed concavity (G), weakly linear (H), and strongly linear (I) relationships.

4.4.2 Pinferna predictions in cell lines and tissues relative to competing alternatives

To be useful for new samples, gene-specific model predictions should be more accurate than guesses based on past copy-number estimates of the protein in other settings. Therefore, we assessed Pinferna predictions against a null model built by iteratively drawing randomized measurements for each protein's abundance from the meta-assembled dataset originally used for training (Figure 4.1A). Accuracy was quantified by subtracting the measured value from the predicted value for each protein, taking the absolute value, and dividing by the standard deviation of the individual protein abundance across the 369 cell lines in the training data. This variance-scaled residual inversely weighs error by the breadth of abundances observed in other biological contexts. Finally, we compared the distribution of variance-scaled residuals between Pinferna and 100 null models to arrive at a proteome-wide estimate of model performance.

The first accuracy test was performed with HeLa cells, a line excluded from one of the original meta-assembled resources. We leveraged an independent study that carefully examined HeLa-to-HeLa differences with paired transcriptomics and SWATH proteomics (38). Pinferna consistently outperformed randomized measurements for all 12 HeLa derivatives investigated ($p < 10^{-6}$; Figure 4.5A and Figure 4.6A). Accuracy estimates were comparable when reported protein abundances were used instead of protein re-quantifications performed exactly as done for the training data (See Materials and Methods; Figure 4.6A and 4.6B). Pinferna predictions were similarly resilient to reductions in transcriptomic sequencing depth—accuracies were comparable down to about 500,000 reads and remained superior to randomized measurements until about 50,000 reads (Figure 4.6C). The results bolster recent claims that typical single-cell RNA-seq data sequenced at ~50,000 reads per cell poorly reflect protein abundances (39, 40) and separately indicate that Pinferna's bulk predictions of protein from mRNA are robust to algorithmic details.

To test Pinferna more broadly against other methods for protein estimation, we integrated the difference between the cumulative distribution functions of a model prediction and the median null to yield a single measure of accuracy improvement (Δ CDF; Figure 4.5B). By Δ CDF, Pinferna was compared with two alternative approaches: 1) PaxDb, a meta-repository of protein abundances largely determined by uncalibrated peptide- and spectral-counting methods (41); and 2) protein-transcript ratios (PTR), which proportionately relate mRNA to protein abundances from data collected in 29 tissues (42). PaxDb and PTR each accommodate generic predictions using all available data and specific predictions

restricted to data from a cell line or tissue of interest; both implementations were tested when possible. For the comparative evaluation, we assembled transcriptomic and SWATH data from 29 cell lines of nine cancers from the NCI-60 panel not included in the meta-assembly (43, 44). Overall, Pinferna was significantly more accurate than randomized measurements, whereas PaxDb and PTR were less accurate (Figure 4.5C). Results were unchanged when transcriptomics were pre-processed with a different alignment pipeline (Figure 4.6D), reinforcing that Pinferna is tolerant of how mRNA TPMs are calculated. We observed no bias in estimates among cancer types (Figure 4.5C) and thus concluded that Pinferna was the preferred method for predicting protein copy numbers from mRNA in cultured cell lines.

The performance in cell lines prompted us to ask whether Pinferna estimates would hold in more complex samples such as tissue. We assembled transcriptomic and SWATH data for 40 primary tumors and 39 normal human prostate samples (45). PaxDb remained significantly worse than randomized measurements ($p < 10^{-14}$; Figure 4.5D and 4.5E), consistent with the recognized limitations of peptide and spectral counts (11, 46). PTR was on par with randomized measurements in prostate cancer samples and significantly better in normal prostate samples, likely due to its tissue-centric focus (42). Nevertheless, Pinferna remained well ahead of all methods, indicating generality of proteome-wide copy number estimates from transcriptomics in primary tumors and nonmalignant tissues.

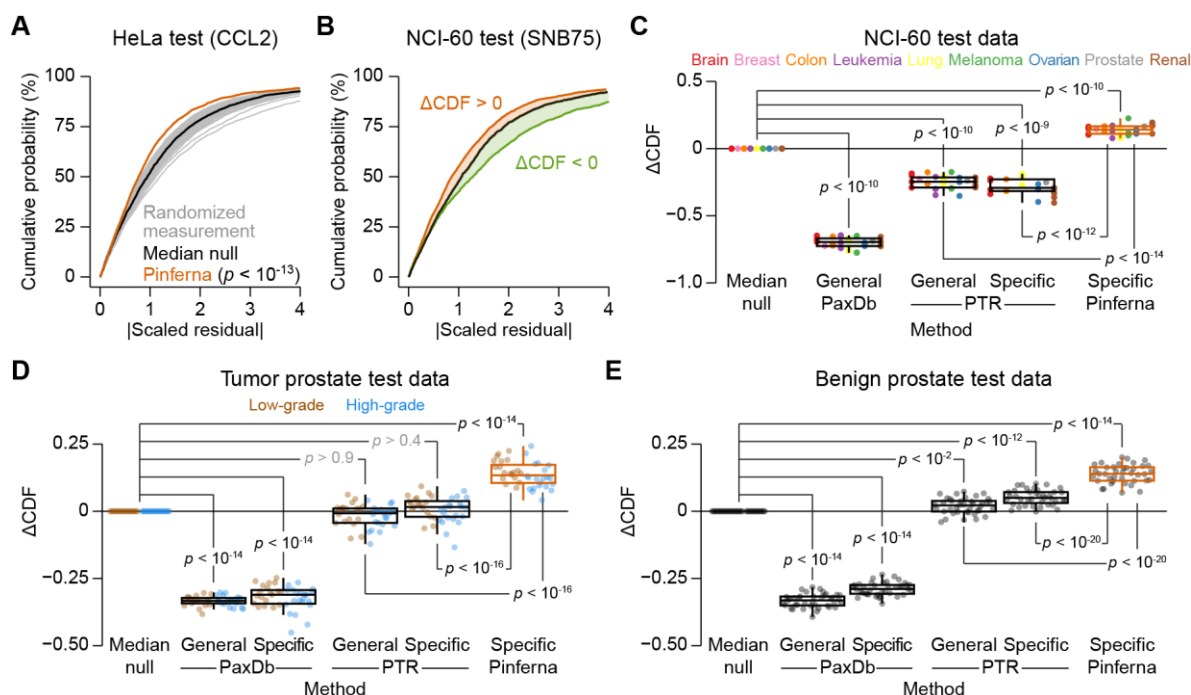


Figure 4.5. Pinferna outperforms empirical guesses and competing methods for absolute protein abundance estimation.

(A) Pinferna compared to random protein-specific guesses. Model predictions were nondimensionalized as a scaled residual by subtracting the measured abundance, dividing by the standard deviation of the SWATH-scaled protein measured across the meta-assembly, and taking the absolute value (|Scaled residual|). The |Scaled residual| cumulative density was compared to randomized measurements drawn from the SWATH-scaled proteomic data for each gene. Randomized measurements were iterated 100 times (gray) to identify a median null (black) that served as a null distribution for model assessment. Left-shifted distributions indicate improved proteome-wide accuracy (relative to each protein's variability) compared to protein-specific randomized measurements. Pinferna predictions of HeLa cells (orange; PRJNA437150; PXD009273) were compared to the null distribution by K-S test ($p < 10^{-13}$).

(B) Aggregate performance assessment of protein abundance predictions. The difference in cumulative density functions between test predictions and the median null distribution (ΔCDF) was integrated to identify approaches that performed better ($\Delta CDF > 0$, orange) or worse ($\Delta CDF < 0$, green) than protein-specific guessing. Data are from a prediction of Pinferna (orange) and tissue-specific protein-to-mRNA ratio (PTR; green).

(C–E) Pinferna is consistently and uniquely superior to empirical guessing. ΔCDF values were calculated for NCI-60 cell lines (C; PRJNA433861; (43)) excluded from model training (Figure 4.1A) and organized by cancer type ($n = 5$ brain, 1 breast, 3 colon, 4 leukemia, 4 lung, 3 melanoma, 3 ovarian, 1 prostate, 5 renal), primary prostate cancer samples organized by grade of the cancer (D; $n = 19$ low-grade, 21 high-grade), and normal prostate tissue (E; $n = 39$; PRJNA579899; PXD004589). PaxDb (41) and PTR (42) were used generically or in a tissue-specific way as alternative approaches (See Materials and Methods). A cell line-specific PaxDb estimate was only available for U251 cells. Differences between groups were assessed by rank-sum test with Šidák correction. Box-and-whisker plots show the median (horizontal line), interquartile range (IQR, box), and an additional 1.5 IQR extension (whiskers) of the data.

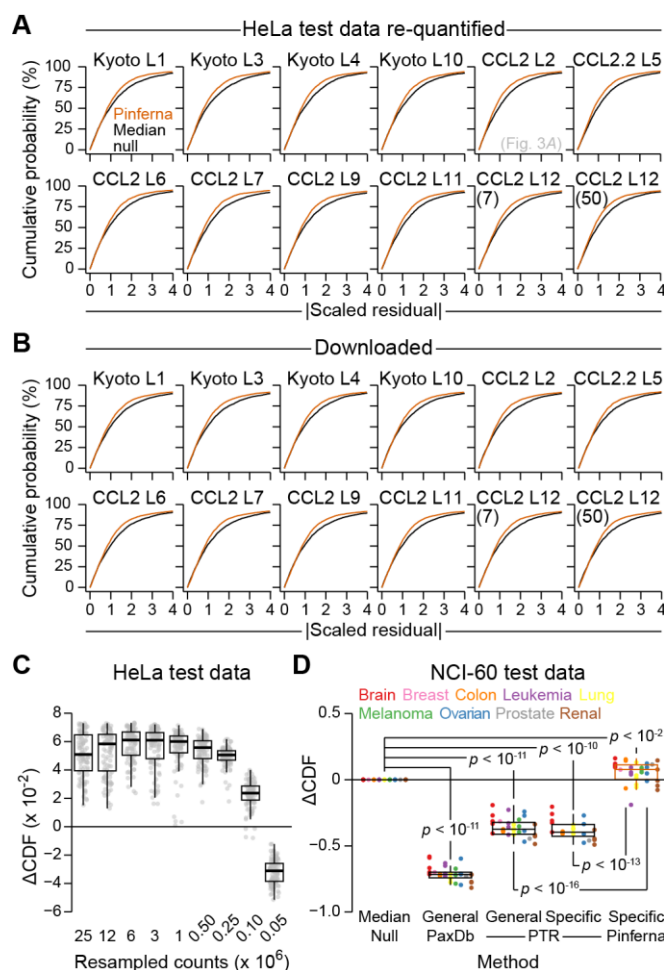


Figure 4.6. Robustness of Pinferna predictions.

(A, B) Prediction accuracy is not dependent on SWATH analytical details. Cumulative distribution plots comparing Pinferna and median-null predictions for paired RNA-seq–SWATH datasets of 12 HeLa derivatives (PRJNA437150; PXD009273) named as in the publication (38). Proteins were re-quantified from raw SWATH data processed exactly like the meta-assembly (See Materials and Methods; A); or, protein quantities were taken directly from the publication (38) (B). Results from the re-quantified CCL2 L2 derivative are reprinted from Figure 4.5A.

(C) Prediction accuracy is not heavily dependent on RNA-seq read depth. Count-based RNA-seq data for the HeLa lines was averaged and iteratively downsampled ($n = 100$ iterations; gray) and TPM values re-estimated before making proteome-wide copy-number predictions by Pinferna. See Figure 4.5B for an explanation of ΔCDF .

(D) Prediction accuracy is not dependent on RNA-seq analytical details. Pinferna predictions were made using TPM values taken directly from the original publication (44), and ΔCDF values were calculated for NCI-60 cell lines excluded from model training (Figure 4.1A) and organized by cancer type ($n = 5$ brain, 1 breast, 3 colon, 4 leukemia, 4 lung, 3 melanoma, 3 ovarian, 1 prostate, 5 renal). Differences between groups were assessed by rank-sum test with Šidák correction.

For (C and D), box-and-whisker plots show the median (horizontal line), interquartile range (IQR; box), and an additional 1.5 IQR extension (whiskers) of the data.

4.4.3 Application to in silico modeling

RNA-seq often substitutes for protein when parameterizing systems-biology models of signaling, metabolism, and cell fate (8–10). For example, in constructing a mass-action model (25) of cardiomyocyte infection by coxsackievirus B3 (CVB3), RNA-seq was used to estimate abundances of the serial CVB3 receptors, CD55 and CXADR (Figure 4.7A). Both estimates were HL extrapolations from a very-limited set of SWATH–RNA-seq pairings, which motivated a direct assessment of protein abundance by quantitative immunoblotting with recombinant standards (25, 47). Direct protein estimation was feasible for cultured cell lines but would be impossible for human hearts, where the severity of CVB3 infections is highly variable (48). We reasoned that Pinferna could provide extensibility to address this challenge and similar needs in cancer (8–10) and neurologic disease (19).

RNA-seq data was collected for 1489 healthy and failing human heart samples from the U.S. and Europe, available through GTEx (49), MAGNet (50), and EGA (51). RNA-seq reads from the three studies were realigned and assembled together to enable comparability (See Materials and Methods; Dataset EV8). The realignment confirmed no biases in expression based on data source for *CD55* (Figure 4.8A). *CXADR* expression increases in cases of cardiomyopathy (52, 53), and we reproduced this result by stratifying cases from the three sources ($p < 10^{-244}$; Figure 4.8B). For both genes, the range of expression in heart samples fell within the variation observed across cancer cell lines in the meta-assembly (Figure 4.8C and 4.8D). *CD55* is an HL+LASSO gene whose predictions are conditionally dependent on nine other genes. These features reduce *CD55* inferences below the smoothed average of the training at very low TPM (Figure 4.8C). By contrast, *CXADR* is an HL gene that is steeply nonlinear at low TPMs where small changes greatly influence the protein copy number estimate (Figure 4.8D). Because of the nonlinear inferences, *CD55*–*CXADR* proteins were strongly coupled even while *CD55*–*CXADR* mRNAs were much less so ($p < 10^{-12}$; Figure 4.8E and 4.8F). Using the Pinferna estimates of *CD55* and *CXADR* as initial conditions, we created a series of individualized model variants to create a virtual cohort of the human population. The individualized models were initiated with a high titer of CVB3 that guaranteed infection of permissive cells, and the concentration of virions was collected at each time point for 24 hours of simulated infection (See Materials and Methods). The goal was to investigate whether inferred *CD55*–*CXADR* protein variations yielded a wide enough range of infection outcomes in the model that one or both receptors could be nominated as a susceptibility factor.

Examining the predicted distribution of viral loads over time, we noted a strong asymmetry in onset of infection (Figure 4.7B). At 12 hours, 64% of individuals were detectably infected, producing mature virions above one plaque forming unit (1 pfu = 0.48

± 0.12 nM in these simulations). By 24 hours, the models yielded a left-skewed distribution, which straddled the mean lytic yield of viruses in the CVB3 genus (~ 100 pfu = 48 ± 12 nM for a $3700 \mu\text{m}^3$ cell) (24, 25, 54). Even at this uncharacteristically late time, 15% of individuals remained uninfected, suggesting they were intrinsically resistant. The remaining cases were best fit as a three-component Gaussian mixture of low, medium, and high susceptibilities (Figure 4.7C). Based on mean lytic yield, we interpreted these groups as prone to subinfection, infection, and severe infection, with failing hearts falling almost entirely into the infection and severe-infection groups (Figure 4.8G). For comparison, we abandoned Pinferna and attempted a randomized-measurement approach by linearly scaling the RNA-seq data about a quantity of CD55 and CXADR arbitrarily selected from the training data (See Materials and Methods). As expected, model outputs were so dependent on the randomized measurement that they were uninterpretable when viewed in aggregate (Figure 4.8H). Randomized measurements tended to predict $\sim 100\%$ resistance or $\sim 100\%$ lytic infections and underestimate the low-susceptibility group, although some fortuitously matched the true inferences. We concluded that the Pinferna-derived model outputs were compelling enough to interpret further.

Among heart samples, the distributions of *CD55* and *CXADR* RNA transcripts were quite different (Figure 4.7D and 4.7E). The range of *CD55* expression was \sim tenfold that of *CXADR*, hinting that it might be the dominant receptor for *in silico* susceptibility. However, these population-wide trends changed when viewed as protein inferences (Figure 4.7F and 4.7G). Both *CD55* and *CXADR* were more symmetrically distributed, with *CXADR* exhibiting greater overall variance. Importantly, when individuals were classified on the basis of their inferred susceptibility, we found that *CXADR* abundance alone was sufficient to stratify the population. This application of Pinferna illustrates how direct substitution of transcriptomics can misconstrue the outputs of systems-biology models built for protein networks.

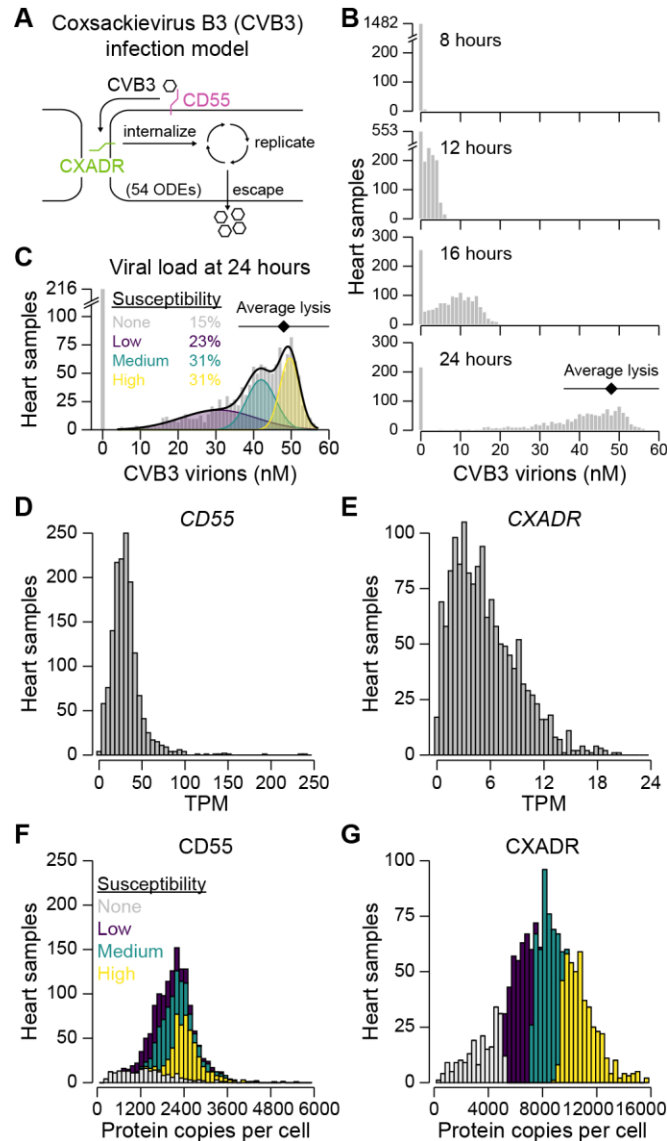


Figure 4.7. Simulating degrees of human cardio-susceptibility to coxsackievirus B3 (CVB3) infection based on inferred abundance differences in CVB3 receptors.

(A) An in silico model of CVB3 initiated by its receptors CD55 and CXADR. After binding, the virus undergoes internalization, replication, and escape. The viral life cycle is mathematically modeled with 54 ordinary differential equations (ODEs; MODEL2110250001).

(B) Distribution of viral load over time from 1489 human heart samples. Inferred abundances of CD55 and CXADR from each sample were used to simulate CVB3 infection. Each model run consisted of 100 simulated infections up to 24 hours with a coefficient of variation in model parameters of 5%. Viral loads (gray) at the indicated time points are shown along with the estimated point of lysis (black: mean estimated lytic yield \pm s.d. (24, 25, 54)).

(C) Four modes of infection susceptibility to terminal CVB3 infection. Viral load at 24 hours was replotted from B fit to a Gaussian mixture model (black) of three components (purple, green, yellow). Relative population densities in each of the susceptibility groups is shown along with the estimated point of lysis (black: mean estimated lytic yield \pm s.d. (24, 25, 54)).

(D, E) Distribution of mRNA abundances for CD55 (D) and CXADR (E) normalized as TPM.

(F, G) Distribution of inferred protein copy numbers per cell for CD55 (F) and CXADR (G) with each sample colored by its susceptibility.

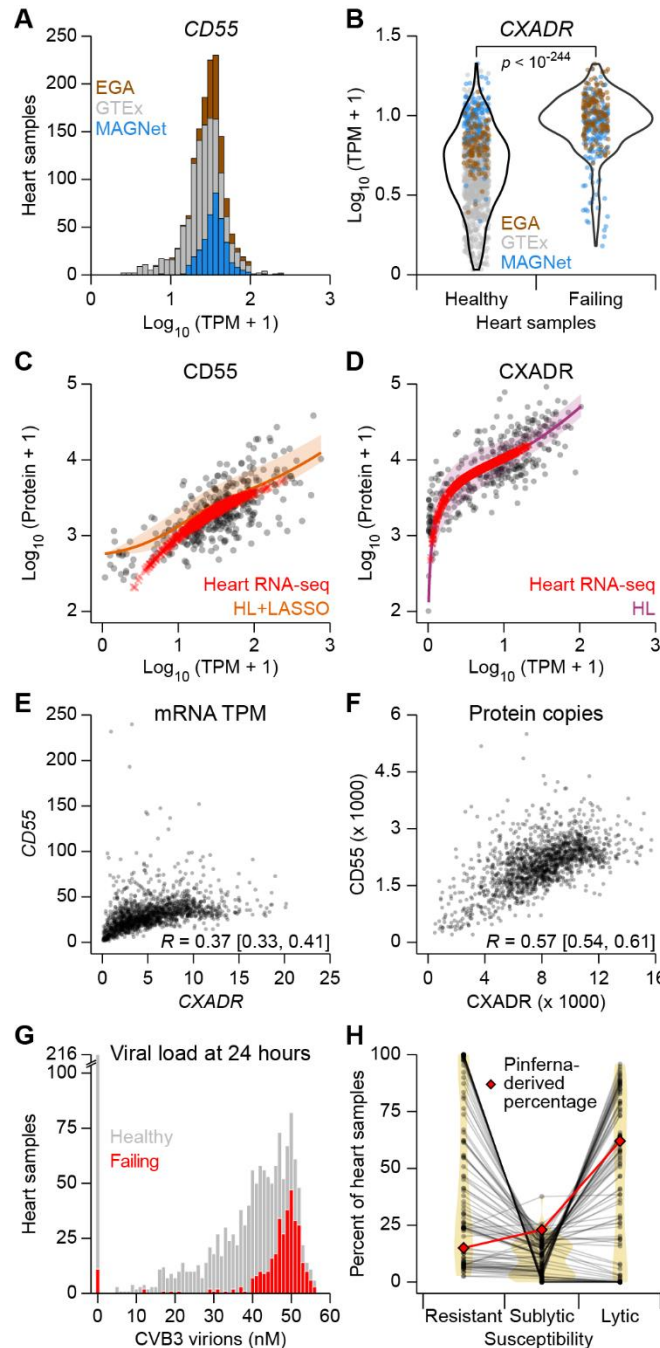


Figure 4.8. Calibrated protein inferences of CD55 and CXADR yield disease-related predictions of coxsackievirus B3 (CVB3) susceptibility.

(A) Distribution of *CD55* abundance (Dataset EV8) separated by source data: EGA (brown; EGAS00001002454), GTEx (gray; phs000424.v9.p2), MAGNet (blue; GSE141910).

(B) *CXADR* is upregulated in failing hearts. *CXADR* abundance (Dataset EV8) was stratified by heart health and colored by source data as in A. Differences between groups were assessed by rank-sum test.

(C, D) Calibration plots (C, orange; D, purple) and Pinferna predictions for CD55 (C) and CXADR (D) in human heart samples (Heart RNA-seq; red). Best-fit calibrations \pm 95% confidence intervals are overlaid on the proteomic–transcriptomic data from $n = 369$ cancer cell lines. CD55 deviations from the smoothed best fit are caused by cardiac-specific features in the HL+LASSO regressions.

(E, F) CD55–CXADR coregulation is increased at the protein level (F) compared to the mRNA level (E). For each comparison, the Pearson R is shown with 95% confidence interval in brackets calculated by the Fisher Z transformation.

(G) Replotted histogram of Figure 4.7C separated by heart health (Dataset EV8).

(H) Predicted prevalence of susceptibility groups based on randomized measurements. After linearly scaling to randomized abundances of CD55 and CXADR ($n = 100$ randomizations), CVB3 infections were simulated for 1489 heart samples as in Figure 4.7B. The 24-hour end states were quantified by the percentage of samples with resistant ([CVB3 virions] = 0 nM), sublytic ($0 \text{ nM} < [\text{CVB3 virions}] < 36 \text{ nM}$), and lytic ([CVB3 virions] $\geq 36 \text{ nM}$) phenotypes (Figure 4.7B). Pinferna-derived percentages are overlaid in red. Results from the randomized simulations are connected, and densities in each group are shown by a violin plot in the background (yellow).

For (B–F), $n = 1489$ heart samples.

4.4.4 Application to molecular subtyping

Transcriptomic profiles are widely used to define disease subtypes (55–57), which may change when gene expression is replaced by inferred protein abundance as a closer surrogate of cell function. As a longstanding example, we selected the intrinsic molecular subtypes of breast cancer defined by a 50-gene classifier (PAM50) for 796 cases with RNA-seq in The Cancer Genome Atlas (26, 58, 59). For consistency, our analysis focused on the 4366 transcripts compatible with protein inference (Figure 4.1A), but results were unchanged when using the entire available transcriptome (Figure 4.9A–D). Consensus clustering of mRNA profiles identified five ordered and stable groups, which were statistically enriched in PAM50-assigned cases of 1) Normal-like, 2) HER2+, 3) Luminal A, 4) Basal-like, and 5) Luminal B breast cancer ($p < 0.004$ by hypergeometric test; Figure 4.10A, 4.9A, and 4.9B). When the analysis was repeated with Pinferna estimates after standardization, the smallest number of stable and significant consensus clusters was again five (Figure 4.9E and 4.9F). However, the enriched PAM50 assignments were reordered, and $186/796 = 23\%$ of cases changed to a different cluster (Figure 4.10A). The aggregate transformations of Pinferna (Figure 4.1C–E) thus exceeded a standardized rescaling and considerably altered subgroup composition.

Among reassigned samples, we noted preferential enrichments in Luminal A (26%) and Luminal B (29%) tumors over Basal-like and HER2+ (Figure 4.10B). Luminal A/B cases are often intermingled in clusters defined by transcriptomics (60), prompting us to look more deeply at their reassignment with Pinferna (Figure 4.9G and 4.9H). We surveyed for HL+LASSO genes whose z-score standardized values changed the most from mRNA to inferred protein and looked within these influential genes for features (other genes) that were STRING interactors (See Materials and Methods; Figure 4.3F and 4.9I). Calibration of the mRNA-to-protein relationship for SDHAF2 (a mitochondrial Complex II assembly factor) was dramatically improved with abundance information from other genes, including its interactors, SDHA and MDH2 (Figure 4.10C). Similarly, inference of SKA3 (a subunit of the mitotic Ska complex) was influenced by multiple binding partners (Figure 4.10D). One of the most notable examples of LASSO modulation was CTNNA2 (an adhesion protein involved in actin regulation). CTNNA2 protein was ubiquitous in the meta-assembly, but its mRNA was undetectable in 23% of training samples; of these, nonzero protein inference of 83% was recovered by using abundances including CTNND1 and plakophilin (PKP2, PKP3) interactors (Figure 4.10E). For cluster reassignments of breast cancer, the gene with the most interaction-rich feature set was CDK4. Like SDHAF2, CDK4 protein abundance was largely independent of its own mRNA, but a useful calibration was achieved when considering various cyclins and other binding proteins (Figure 4.10F). This result is

important because CDK4/6 inhibitors are approved to treat luminal breast cancers (61, 62), but responsiveness has not consistently associated with the abundance of *CDK4* mRNA (63, 64). Newly meaningful subtypes with therapeutic implications might arise when combining transcriptomics with Pinferna to get closer to the functional proteome.

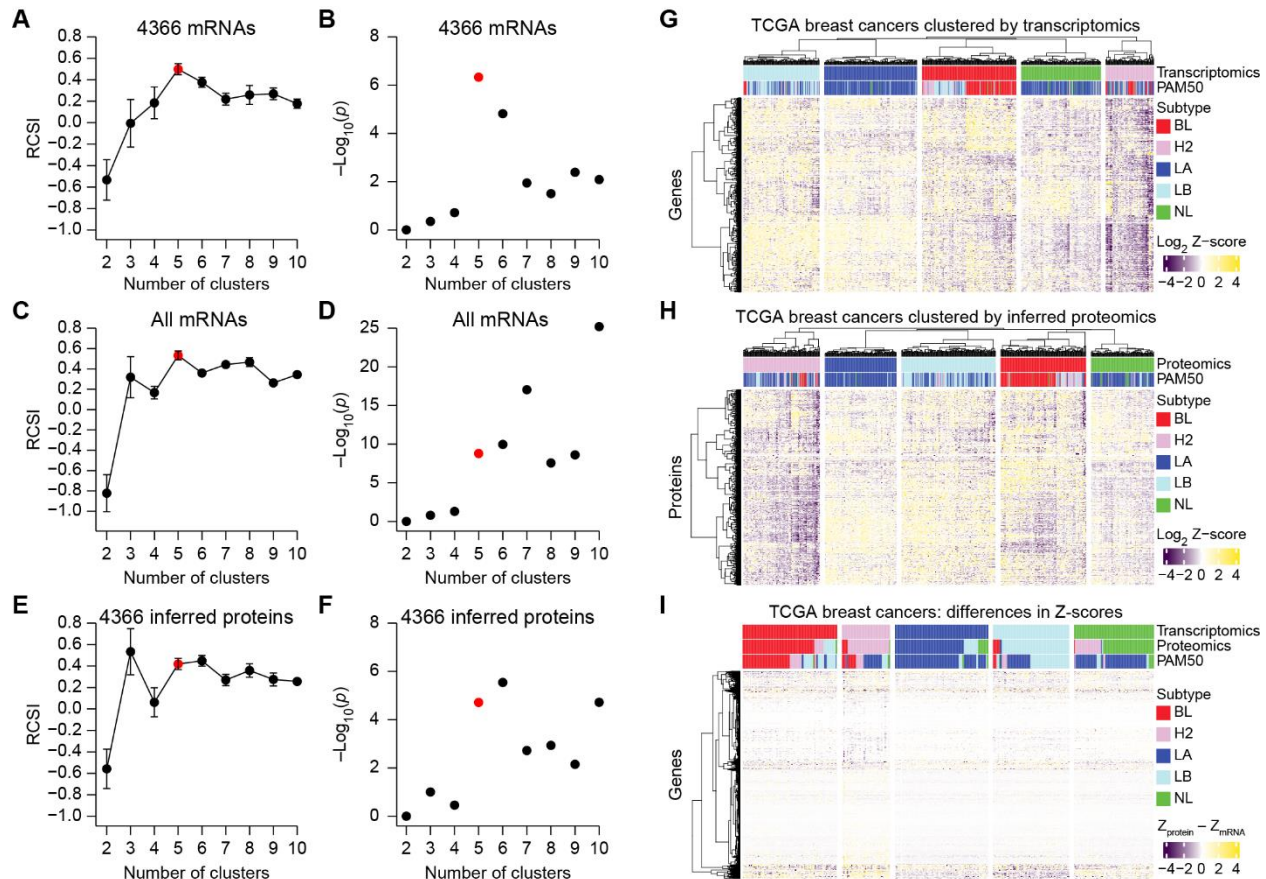


Figure 4.9. Consensus re-clustering of 796 breast cancer cases from The Cancer Genome Atlas (TCGA) by protein inference.

(A–F) Relative cluster stability index (RCSI; A, C, and E) and significance (B, D, and F) of cluster number for consensus clustering (See Materials and Methods) applied to 4366 mRNAs used by Pinferna (A, B), all mRNAs quantified by RNA-seq (C, D), and protein inferences by Pinferna (E, F). Stable maxima or near maxima at five clusters is indicated in red.

(G, H) Two-way hierarchical clustering within transcriptomic (G) and inferred proteomic (H) clusters annotated as in Figure 4.10A.

(I) One-way hierarchical clustering of genes by differences in standardized Z-scores between inferred protein and mRNA ($Z_{\text{protein}} - Z_{\text{mRNA}}$).

For (G–I), hierarchical clustering was performed by Euclidean distance and Ward's linkage. BL: Basal-like; H2: HER2+; LA: Luminal A; LB: Luminal B; NL: Normal-like.

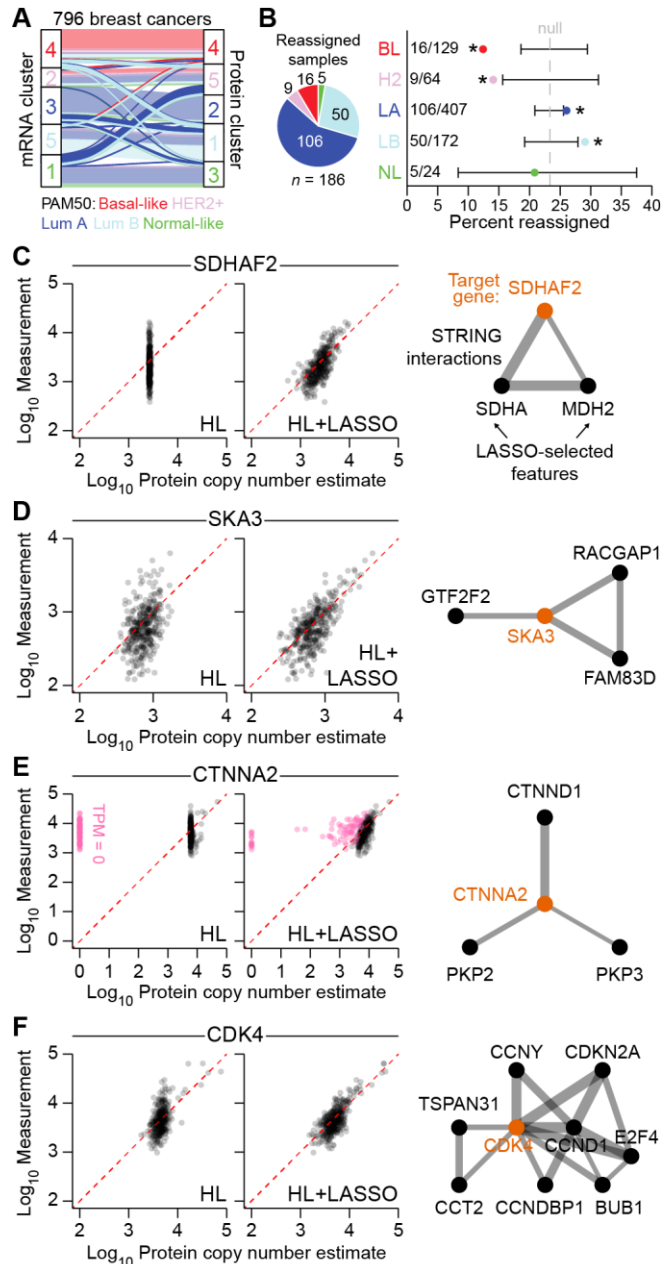


Figure 4.10. Inferred proteomics reassigns luminal A/B transcriptomic subtypes of breast cancer.

(A) Reorganization of five consensus clusters defined by RNA-seq (left) and Pinferna (right) for 796 breast cancers in The Cancer Genome Atlas (26). Clusters were determined by Monte Carlo consensus clustering (65) and colored according to the dominant PAM50 subtype of each cluster. Samples that did not change clusters are transparent in the background while samples that changed are opaque in the foreground. Lum A: Luminal A; Lum B: Luminal B.

(B) Reassigned samples are predominated by luminal A/B PAM50 subtypes. (Left) Proportion of each subtype among samples that were reassigned. (Right) Percent reassignments for each subtype. The average overall reassignment rate is shown as a null reference ($186/796 = 23\%$; gray dashed) with the 90% hypergeometric confidence interval (black) for each subtype. Reassignment enrichments were determined by hypergeometric test, asterisk indicates $p < 0.05$. BL: Basal-like; H2: HER2+; LA: Luminal A; LB: Luminal B; NL: Normal-like.

(C–F) Cluster-reorganizing genes are highly dependent on other genes. (Left) Concordance between SWATH measurements and the HL fit \pm LASSO in the meta-assembly. Perfect concordance is given by the red dashed line. Pink points in E are samples with TPM = 0 for *CTNNA2*. (Right) STRING interactions (edges) among the target gene (orange) and its LASSO-selected features (black). Edge thickness (gray) reflects the confidence of the interaction as determined by STRING. Thicker lines represent a higher confidence score. Line lengths are arbitrary.

4.5 Discussion

We have devised a straightforward gene-by-gene formalism that uses mRNA to achieve absolute protein estimates informed by the best measurements available for each data type. Gene-specific inferences are gleaned from cancer cell lines, but general accuracy and utility is verified in multiple other contexts. Although not exhaustive, our coverage of 4366 mRNA–protein relationships is considerable given that modern SWATH experiments reliably quantify ~5000 proteins (38). Encouraged by the robustness of predictions to read depth and alignment details, we provide Pinferna as an open resource (<http://janeslab.shinyapps.io/Pinferna>; Dataset EV9). The platform is optimal when provided full RNA-seq profiles, but it also accommodates single-gene TPM entries for LASSO-free inference when subset data are exported from public repositories. The hope is to attract general users to the value in seeing genes and gene profiles of interest through the lens of inferred proteomics (source code is available for developers; See Data Availability).

The mRNA-to-protein relationship classes used here distill the major findings of prior models that were more fine grained (42, 66, 67). Measured mRNA is the net result of its transcription–degradation, with the per-mRNA yield of translation for each gene being the greatest determinant of absolute protein abundance (20, 66). Together, mRNA abundance and per-mRNA yield define an expected set point for protein abundance as captured by the HL relationship class. For typical in vitro cultures, cell doubling is faster than turnover of most proteins (17, 66), creating a perpetual state of halving-and-recovery that likely explains why many HL models are log-concave up. Nonetheless, overall accuracy of Pinferna did not decrease with clinical samples that were less proliferative, suggesting a role for other cyclic perturbations in vivo, such as circadian rhythms (68). Some genes additionally require protein complexes to persist stably (17, 21), which creates buffering dependencies on other genes that are coexpressed. The HL+LASSO approach seeks to capture this relationship class by identifying statistical mRNA–protein associations in trans. Despite its heavy L1 regularization (69), LASSO recovered a significant number of documented protein-protein interactions. Recently, a reciprocal approach to predict relative protein abundance from mRNA was proposed that constrains the search space for each gene to its CORUM–STRING interactors but relaxes the regularization by using elastic net (30, 69). These models retain many more features (158–457 (30) compared to 1–83 features for HL+LASSO), but their relative predictions cannot be compared with the absolute copy-number estimates of Pinferna. Future versions may consider hybrid regularizations that penalize CORUM–STRING interactors less during LASSO feature selection. Iterative approaches might also use Pinferna inferences as

LASSO features for other genes to approximate biological dependencies more closely. Lastly, we speculate that M genes arise from protein complexes so large that pairwise interactions within a complex completely dictate abundance (70, 71).

Absolute protein estimates are of great practical use for making qualitative determinations in systems-biology models. For example, using inferred protein to simulate heart infections, we clarified that individuals with fewer than ~5000 copies of CXADR per cell were not susceptible to CVB3 (Figure 4.7G). Another subfield of relevance is genome-scale metabolic modeling with tailored derivatives of the generic human metabolic network reconstruction (72). For cell- or tissue-specific modeling, the generic reconstruction is pruned according to which metabolic genes are “not expressed” in a biological context of interest (73, 74). Irrespective of the pruning algorithm, the choice of threshold is made absolutely across all genes in a sample, which defines the resulting model complexity (75). Protein abundances for some metabolic pathways scale linearly with mRNA, but others do not (6, 7), and protein–mRNA set points vary over several orders of magnitude (20, 42). For metabolic models concerned with protein fidelity, estimating copy numbers from RNA-seq is a scalable alternative to proteome immunohistochemistry (73).

There are limitations in our approach to absolute copy-number estimation. By relying on SWATH for calibration, we lose many of the 8000+ proteins quantified in relative terms by TMT (27). Calibration data were all collected at steady state; thus, we caution against using RNA-seq obtained shortly after acute perturbations when transcripts and proteins will be most uncoupled (17). Signaling proteins rapidly turned over by ubiquitylation (TP53, NFE2L2, NFKBIA) might require other formalisms when they become detectable by SWATH. Broadening Pinferna predictions to non-human samples awaits the availability of robust SWATH libraries in other mammals (76). Last, we recall that no total-protein estimator captures functional state, such as the surface localization of CD40, the tyrosine phosphorylation of PKP2–PKP3, or the kinase activity of CDK4 (Figures 4.3C, 4.10E, and 4.10F). Despite these caveats, the method adds an immediately useful approach for systems biology that compares favorably against existing alternatives.

In statistics, bootstrapping by computation allowed the field to tackle analytical needs that were impossible to address by existing methods (77). Analogously, proteomics can extend its reach by combining with RNA-seq to inform more biological samples, including retrospective ones it would never otherwise have access to. There is already precedent for such “transcriptomic bootstrapping” to infer genomic alterations, molecular kinetics, and cellular trajectories (78–80).

4.6 Materials and Methods

SWATH alignment and quantification

Raw SWATH data files were obtained from the PRIDE repository (15) (CAL51, PXD003278 (28); U2OS, PXD000954 (29); HeLa, PXD00927 (38)) and converted to .mzML format using the MSConverGUI (version 3.0) in the ProteoWizard software suite (81) with the following options: Output format, mzML; Extension, mzML; Binary encoding precision, 64-bit; Write index; Use zlib compression. Peptide fragments were aligned with OpenSwathWorkflow in OpenMS (version 2.4.0) (82) with the following options: -sort_swath_maps, -readOptions normal, -batchSize 1000, -use_ms1_traces, -mz_correction_function quadratic_regression_delta_ppm. Statistical control was performed with PyProphet (version 2.2.5) (83) with the following options: --group_id=transition_group_id, --tric_chromprob. Each series of SWATH runs was realigned with TRIC (84) using msproteomicstools (version 0.11.0) with the following options: --method LocalMST, --realign_method lowess, --max_rt_diff 60, --mst:userRTCorrection True, --mst:Stdev_multiplier 3.0, --target_fdr 0.01, --max_fdr_quality 0.05, --alignment_score 0.0005. The top three peptide fragments by intensity (or all peptide fragments if fewer than three) were summed for each protein to estimate relative abundance. Summed intensities were mean-averaged across technical replicates when available. To place summed intensities on an absolute scale, the median abundance of all detected proteins within each sample was centered at 10,000 protein copies per cell (38).

RNA-seq alignment and quantification

For all studies other than Figure 4.7, SRA files were obtained from the Sequence Read Archive (SRA) (85) (HeLa, PRJNA437150; NCI-60, PRJNA433861; Prostate, PRJNA579899) and converted to raw FASTQ files using sratoolkit (version 2.10.5) with fasterq_dump. TruSeq adapters were trimmed using the fastq-mcf function in the ea-utils package with the following options: -q 10, -t 0.01, -k 0. Trimmed datasets were aligned to the human genome (GRCh38) using HISAT2 (version 2.1.0) (86) with the following options: --dta (downstream transcriptome assembly) and either --rna-strandedness RF (for paired-end reads generated by the TruSeq strand-specific library; NCI-60 and prostate samples) or --rg-id (for single-end reads generated by the TruSeq library; HeLa). Output SAM files were converted to BAM files using the sort function in samtools (version 1.12) (87), and BAM files were indexed to create BAI files using the index function for obtaining counts downstream. Alignments were assembled into transcripts using StringTie (version 2.1.0) (88) with the -e option restricting assembly to known transcripts in the provided annotation.

Counts were obtained using HTSeq (version 2.0.2) (89) using BAM files as the input with the following options: -f bam, -r pos, -m intersection-strict, -s reverse, -a 1, -t exon, -i gene_id. Heart RNA-seq data were obtained from dbGaP (GTEx, phs000424.v9.p2), the Sequence Read Archive (MAGNet, SRP237337), and the European Genome-Phenome Archive (EGA, EGAS00001002454). GTEx samples were converted from BAM files to raw FASTQ files using samtools (version 1.12) (87). MAGNet samples were converted to raw FASTQ files using sratoolkit (version 2.10.5) (<https://hpc.nih.gov/apps/sratoolkit.html#doc>). EGA samples were converted to raw FASTQ files using samtools (version 1.10/1.12) (87) and paired reads were recreated using the fastqCombinePairedEnd.py script from Eric Normandeau (<https://github.com/enormandeau/>). Datasets were aligned to the human genome (GRCh38) using HISAT2 (version 2.1.0) (86) with the --dta option for downstream transcriptome assembly. Output SAM files were converted to BAM files using samtools (version 1.12) (87). Alignments were assembled into transcripts using StringTie (version 2.1.0) (88) with the -e option restricting assembly to a unified list of transcripts that was provided by first running StringTie using the --merge option.

Data harmonization

The table of MANE Select identifiers was obtained from the source publication (90) and filtered for “MANE Select” genes. The filtered table was appended with UniProt accession codes using biomaRt (version 2.52.0) and GRCh38. The Ensembl BioMart browser was used to obtain HGNC identifiers, Ensembl transcript identifiers (with version numbers for maximum overlap), RefSeq mRNA identifiers, NCBI (formerly Entrez) gene identifiers, UniProt accession codes, and UniProt gene symbols for *Homo sapiens*. Each row of the MANE Select table was matched to at least two identifiers in the biomaRt table to determine the UniProt accession numbers. When MANE Select annotated a gene symbol as LOC##### and biomaRt contained a more descriptive gene symbol, the biomaRt gene symbol replaced the MANE Select gene symbol and the “Database” column was updated to include “biomaRt symbol” as the source. The harmonization identified 83 genes that are not currently available in UniProt. The final harmonized table of ten identifiers (nine for the 83 genes not in UniProt) for 19,062 genes is available in Dataset EV2.

Cancer Cell Line Encyclopedia pre-processing

The TMT proteomic dataset was obtained from the source publication (27) as a CSV file (protein_quant_current_normalized.csv). After removing proteins annotated as “Fragments”, gene symbols were matched to UniProt accession codes by using the

harmonized identifier table (Dataset EV2). Protein isoforms with redundant gene symbols were summed. The RNA-seq dataset (31) was obtained from the Depmap portal, back-transformed from \log_2 to TPM, and renamed with the harmonized identifier table (Dataset EV2).

Meta-assembly, calibration, and inference

1. Scaling. For each gene, SWATH copy-number estimates were divided by the corresponding harmonized TMT data for U2OS and CAL51 cells to calculate U2OS- and CAL51-specific scaling factors. Scaling factors were averaged when possible; otherwise, a single scaling factor was used (Figure 4.2A). The resultant scaling factors were then multiplied across the harmonized TMT data table to yield a SWATH-scaled proteomics dataset of 4385 total proteins across 375 cell lines (Dataset EV1).

2. Regression. The SWATH-scaled proteomics and RNA-seq transcriptomics datasets were filtered before regression. As recommended (91), proteomics data from replicates of CAL120 (CAL120_BREAST_TenPx02), SW948 (SW948_LARGE_INTESTINE_TenPx11), and HCT15 (HCT15_LARGE_INTESTINE_TenPx30) were excluded. RNA-seq data were filtered to include only cell lines with SWATH-scaled proteomics available. Both datasets were filtered to retain genes for which SWATH-scaled proteomics was available in at least 150 cell lines (4445/4513 = 98.5% of all SWATH-scaled proteins). Filtered datasets are available in Datasets EV3 and EV4. Numerical approaches for building the M, HL, and HL+LASSO models and assessment of confidence intervals is described in Appendix Supplementary Methods.

3. Model selection. The BIC for each regression was calculated under the assumption of normally distributed random errors as follows:

$$BIC = p \cdot \log(n) - 2 \cdot \log \left(\sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-(y_i - \hat{y}_i)^2 / 2\sigma^2} \right) \text{ where } \sigma = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

where σ is the standard deviation of the fit, n is the number of observations, p is the number of model parameters, \hat{y}_i is the predicted value of the i^{th} observation, y_i is the measured value of the i^{th} observation, and LL is the log-likelihood with \log being the natural logarithm. Comparison of HL with linear, hyperbolic, and 3-parameter logistic alternatives is described in Appendix Supplementary Methods.

Gene ontology analysis

Enrichments of M, HL, and HL+LASSO genes for biological processes were evaluated with the GO knowledgebase (92).

Concavity analysis

The concavity of HL fits was assessed with the `check_curve` function of the `inflection` package (version 1.3.6) in R. HL curves for TPM > 5 (~1 copy per cell) were analyzed after log transformation of x and y coordinates.

Feature weight distributions

To obtain LASSO feature weights, each LASSO coefficient for a gene was multiplied by the mean TPM of the feature averaged across the 369 cell lines in the meta-assembly (including zeros). Feature weights for all HL+LASSO genes were concatenated and filtered for subunits of the proteasome (*PSM*-prefixed gene names) or ribosome (*RPS*- or *RPL*-prefixed gene names), allowing duplicates if the feature appeared in more than one gene. Distributions were plotted as smoothed densities with the `geom_density` function in `ggplot2` (version 3.4.0).

STRING interactions and maps

STRING interactions were obtained with the `STRINGdb` (version 2.8.4) and `rbioapi` (version 0.7.7) packages in R. Sessions were initialized with `STRINGdb$new` and the following arguments: `species = 9606` (Homo sapiens), `version = 11.5`, `score_threshold = 400` (medium-confidence interactions). HL+LASSO genes were mapped with the `string_db$map` function, and up to 1000 medium-confidence interactions were retrieved with the `rba_string_interaction_partners` function and harmonized with the gene identifier table (Dataset EV2). For comparison, LASSO-selected features were substituted with an identical number of randomly selected genes to recalculate the number of interactions. The substitution–recalculation step was iterated 10^4 times to build a null distribution. Details about STRING visualizations are described in Appendix Supplementary Methods.

Assembly of test datasets

HeLa. Using the latest analytical procedures, SWATH data from two HeLa derivatives [Kyoto L8 and CCL2 L13] in the original study (38) did not pass the internal calibration step of the OpenSwathWorkflow alignment and were omitted here. Pre-quantified SWATH data for HeLa derivatives were downloaded from <https://helaprot.shinyapps.io/crosslab/> and normalized to the median copy number of

proteins co-quantified in the meta-assembly (10,000 copies per cell). Downsampling of HeLa RNA-seq reads is described in Appendix Supplementary Methods.

NCI-60. Pre-aligned SWATH data for the NCI-60 panel of cell lines (43) were downloaded from CellMiner as a processed data set (Protein: SWATH (Mass spectrometry) - Peptide), quantified for protein, harmonized as described above, and normalized to the median copy number of proteins co-quantified in the meta-assembly (14,000 copies per cell). Pre-aligned RNA-seq data for the NCI-60 panel of cell lines (44) were downloaded from CellMiner as a processed data set (RNA: RNA-seq - composite expression), summarized as TPM, and harmonized as described above. The following NCI-60 lines excluded from the meta-assembly were used as test data: SF268 (brain), SF539 (brain), SNB-19 (brain), SNB-75 (brain), U251 (brain), Hs 578T (breast), COLO 205 (colon), HCC2998 (colon), KM12 (colon), CCRF-CEM (leukemia), HL-60(TB) (leukemia), MOLT-4 (leukemia), SR (leukemia), EKVX (non-small cell lung), HOP-62 (non-small cell lung), HOP-92 (non-small cell lung), NCI-H322M (non-small cell lung), M14 (melanoma), Malme-3M (melanoma), MDA-MB-435 (melanoma), NCI-ADR-RES (ovarian), OV-CAR5 (ovarian), SK-OV-3 (ovarian), DU145 (prostate), ACHN (renal), RXF 393 (renal), SN12C (renal), TK-10 (renal), and UO-31 (renal).

Prostate. Pre-aligned SWATH data for normal and malignant prostate (45) were downloaded from PRIDE (PXD004589), quantified for protein, harmonized for gene names as described above, and normalized to the median copy number of proteins co-quantified in the meta-assembly (14,000 copies per cell). Harmonization of samples was more challenging because of different patient-coding schemes for the SWATH and RNA-seq datasets. We obtained metadata annotation for RNA-seq from the SRA Run Selector (PRJNA579899) and then reconciled these identifiers with the PXD004589 identifiers using a key personally communicated by Wenguang Shao (93). Tumor-normal pairs were retained in the harmonized dataset if the tumor grade annotations were consistent between SWATH and RNA-seq. The final patient annotations and cross-referencing key is available in Dataset EV7.

Heart. The GTEx test dataset from the v8 final data release consisted of 432 left ventricle and 429 atrial appendage autopsy samples from 561 healthy donors (49). Both GTEx heart tissue sites were considered separately in the analysis. The MAGNet test dataset consisted of 200 cardiomyopathy and 166 healthy control samples (50). The EGA test dataset consisted of 149 cardiomyopathy and 113 healthy control samples (51). After RNA-seq alignment as described above, the 1489 samples were concatenated without batch correction before the analysis.

Breast cancer. Pre-aligned RNA-seq data for ductal and lobular neoplasms in The Cancer Genome Atlas (TCGA) were downloaded from the Genomic Data Commons portal.

The samples were intersected by TCGA identifiers with the published samples profiled by RNA-seq and classified by PAM50 (26), yielding 796 samples in the test dataset.

Alternative methods for protein abundance estimation

PaxDb. The following aggregated proteomics data were downloaded from PaxDb (41) as averaged protein parts per million (ppm): for NCI-60 comparisons, H.sapiens - Cell line (Integrated); for general tissue comparisons, H.sapiens - Whole organism (Integrated); for prostate tissue comparisons, H.sapiens - Prostate gland (Integrated). PaxDb entries less than 0.01 ppm were excluded, and the filtered data were harmonized as described above. Last, each aggregated dataset was normalized to the median copy number of proteins co-quantified in the meta-assembly: for Cell line (Integrated), 8000 copies per cell; for Whole organism (Integrated), 8000 copies per cell; for Prostate gland (Integrated), 9000 copies per cell. Note that PaxDb does not use information from RNA-seq and thus makes a single prediction of protein abundance for each integrated context. Cell line-specific information was not available for NCI-60 lines other than U251.

PTR. Protein-to-mRNA ratios in the original publication (42) were not calculated on the scale of copies per cell. To convert, the protein abundances used for PTR estimation were normalized to the median copy number of proteins co-quantified in the meta-assembly (9000 copies per cell). Using the renormalized protein abundances, we rederived PTRs from the associated RNA-seq fragments per kilobase per million mapped reads (FPKM) as follows: $\text{PTR} = \log_{10}(\text{protein abundance}) - \log_{10}(\text{FPKM})$ (42). PTRs were calculated for each of 29 tissues and in a general manner by using the median PTR. NCI-60 predictions used the PTR specific to each cell line's tissue of origin, which was not available for leukemic, breast, or melanoma lines; therefore, these lines were omitted from PTR predictions.

Randomized measurements, median null model, and ΔCDF

Randomized SWATH measurement profiles were constructed by randomly sampling a gene-specific copy number estimate for each gene in the proteome and iterating 100 times without replacement. The 100 randomized-measurement distributions were compared to Pinferna and ordered by the K-S statistic, with the median distribution selected as the null model for formal K-S hypothesis testing. To compare prediction methods to the median null, the area between the two was integrated by difference (ΔCDF) using the AUC function in the DescTools (version 0.99.47) package in R.

Monte Carlo consensus clustering (M3C)

Consensus clustering of the breast cancer dataset was performed with all genes (19,062) or all genes inferable by Pinferna (4366). LASSO-modulated protein inferences below zero copies per cell were set to zero. All data were log₂ transformed, and genes with zero variance were eliminated before Z-score standardization. Datasets were clustered using the M3C function from the M3C (version 1.18.0) package with the following options: removeplots = T, iters = 10, objective = “entropy”, clusteralg = “spectral”. Clustering statistics lie within the M3C object, and five consensus clusters were selected based on maximum or near-maximum cluster stability and significance. Other hierarchical clustering of the breast cancer dataset is described Appendix Supplementary Methods.

Transcription-translation model

The system of ordinary differential equations was solved in MATLAB R2022a using ode15s. The simulation was performed 100 times allowing parameters to vary lognormally about their central nondimensionalized estimate with a coefficient of variation of 10%. After each simulation, 10 time points were chosen randomly and stored, for a total of 1000 points at the end of the simulation.

CVB3 model

A mass-action model of CVB3 infection (MODEL2110250001) was modified from its published version (25) to accommodate CD55 and CXADR abundances as input parameters. Each simulated infection was initialized with abundances for CD55–CXADR and 10 plaque-forming units of CVB3. In silico infections proceeded for 24 hours and were iterated 100 times with 5% lognormal coefficient of variation between runs. The median virion output was stored, and overall viral load (measured as mature CVB3 virions and after excluding cases of zero viral load) was fit with a Gaussian mixture model using the Mclust function in the Mclust (version 6.0.0) package in R. The best mixture model by BIC was a three-component model of unequal variance, which classified each sample based on the probability of the sample falling into that component. For the randomized-measurement case (Figure 4.8H), protein estimates were obtained by setting the median TPM value of the heart samples to a randomly selected protein set point in the meta-assembly and linearly scaling the other samples around that point. This process was replicated 100 times to create a 1489 sample x 100 replicate matrix of randomized measurements for CD55 and CXADR. Each CD55–CXADR pair was passed to the CVB3 model and simulated with MATLAB (version R2022b) as before 100 times each for a total of 14,890,000 simulations, which were threaded to 100 cores over 10 nodes on the Rivanna high-performance

computing cluster of the University of Virginia. Taking viral load at 24 hours of infection as the phenotype, we classified [CVB3 virions] = 0 nM as resistant, $0 \text{ nM} < [\text{CVB3 virions}] < 36 \text{ nM}$ as sublytic, and $[\text{CVB3 virions}] \geq 36 \text{ nM}$ as lytic.

4.7 Appendix Supplementary Methods

Median (M) regression

For each gene, the median was calculated across the SWATH-scaled proteomic dataset, taking empty entries as missing elements rather than zeros. A 95% confidence interval of the median was estimated by bootstrapping ($n = 1000$ runs).

Hyperbolic-to-linear (HL) regression

For each gene, an HL model was constructed as follows:

$$\text{SWATH-scaled protein copies per cell} = a \cdot \left(\frac{b \cdot \text{TPM}}{c + \text{TPM}} + \text{TPM} \right),$$

where a , b , and c were regression coefficients estimated by nonlinear least squares in MATLAB (version R2022a). To prevent discontinuities from division by zero, c was constrained to be greater than zero. Additionally, a logistic weighting of the cost function was desired to prevent high-abundance cell lines from overleveraging the regression. To achieve both, we devised a two-step procedure in which initial estimates were made with a linear cost function and $c > 0$ constraint using `lsqcurvefit` with 'FiniteDifferenceType' set to 'Central'. The regression estimates of a , b , and c were then used as an initial guess for `nlinfit` with 'RobustWgtFun' set to 'logistic', and the regression was repeated without constraints. If the updated value of c was less than zero, then the updated regression estimates of a and b were used as initial guesses for a second round of `lsqcurvefit` with a linear cost function and $c > 0$ constraint. A 95% confidence interval of the HL fit was estimated by asymptotic error analysis of the regression coefficients using the F distribution to describe the ratio of the sum-of-squared errors for the ideal and parameter-perturbed model divided by their corresponding degrees of freedom.

HL + least absolute shrinkage and selection operator (HL+LASSO) regression

Residuals from the HL fit were regressed against all other genes in the transcriptome by LASSO with the `glmnet` package (version 4.1-6) in R. To determine the optimal penalty strength parameter (λ) for each LASSO regression, we used the `cv.glmnet`

function for cross-validation after increasing the function's minimum fractional change in deviance for stopping (fdev) to 0.01. We accounted for differences among cross-validation runs by iterating cv.glmnet 100 times, calculating the BIC for the best λ in each iteration, and defining the best λ with the lowest BIC as optimal. The optimal λ was used with the glmnet function and all observations of the gene to obtain a regularized feature set and linear coefficients. Output of the LASSO regression was subtracted from HL fit to obtain the HL+LASSO model. LASSO displacements were propagated linearly to the 95% confidence interval of the HL fit, and graphical displays were LOESS smoothed with the geom_smooth function in ggplot2 (version 3.4.0) in R.

Comparison of HL model to alternatives

For each gene, three alternative models were constructed as follows:

$$\text{Linear: } \textit{SWATH-scaled protein copies per cell} = a \cdot \textit{TPM}$$

$$\text{Hyperbolic: } \textit{SWATH-scaled protein copies per cell} = \frac{a \cdot \textit{TPM}}{b + \textit{TPM}}; b > 0$$

$$\text{3-parameter logistic: } \textit{SWATH-scaled protein copies per cell} = a - \frac{a}{1 + (\frac{\textit{TPM}}{b})^c}; a > 0, b > 0$$

where a , b , and c were regression coefficients to be estimated in MATLAB (version R2022a). Linear models were regressed using fitlm with 'RobustOpts' set to 'logistic' and 'Intercept' set to false. Hyperbolic and 3-parameter logistic models were fit similarly to HL but with a combination of lsqcurvefit for constrained regression and fitnlm (which calls nlinfit internally) for logistic weighting of the cost function and the estimation of log-likelihood, which was used for Bayesian Information Criterion (BIC) calculation with the aicbic function. BIC weights (BICw) were calculated as follows:

$$BICw_i = \frac{\exp\left[-\frac{1}{2}(BIC_i - BIC_{min})\right]}{\sum_{k=1}^K \exp\left[-\frac{1}{2}(BIC_k - BIC_{min})\right]}$$

where BIC_i is the BIC for the i^{th} model and BIC_{min} is the minimum BIC in the group of models (34).

STRING visualization

Interaction maps were drafted on the STRING database web site (36) under Search>Multiple proteins. The default output maps were altered as follows: meaning of

network edges = confidence, minimum required interaction score = medium confidence (0.400), disable 3D bubble design, and disable structure previews inside network bubbles. Interaction maps were exported as vectorized SVG files for further stylistic refinement.

HeLa RNA-seq downsampling

Before downsampling, RNA-seq data from the HeLa derivatives were averaged. Raw counts were converted to counts per million (CPM), and the counts per million-normalized reads for each gene were averaged across all derivatives. Then, the average CPM was converted back to an averaged count by multiplying the average read depth and rounding to the nearest integer. The average counts for each gene were downsampled 100 times using `rbinom` in R, with the number of trials equal to the downsampled read depth (25 million to 50,000) and the probability of success equal to the number of average counts for that gene divided by the total number of average counts. Downsampled counts were converted to TPM and used with Pinferna to predict the mean-averaged SWATH data from the HeLa derivatives.

Hierarchical clustering

Breast cancer RNA-seq was \log_2 -transformed and row-standardized. For inferred proteomic profiles, M genes were removed before standardization because of zero variance. Data were clustered by Euclidean distance with Ward's linkage using the `Heatmap` function in the `ComplexHeatmap` (version 2.12.1) package. Columns were clustered within a subtype defined either by transcriptomics or proteomics. To identify genes that changed disproportionately between mRNA and protein, the Z-scores of the species were subtracted: $Z_{\text{diff}} = Z\text{-score}_{\text{Protein}} - Z\text{-score}_{\text{mRNA}}$. To identify genes of interest that drove luminal reassignments, we filtered for genes with a $|Z_{\text{diff}}| \geq 3$ and ranked by the frequency of occurrence, focusing on genes with $|Z_{\text{diff}}| \geq 3$ in six or more samples that had undergone subtype reassignment.

4.8 References

1. R. Phillips, R. Milo, A feeling for the numbers in biology. *Proc Natl Acad Sci U S A* **106**, 21465–21471 (2009).
2. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
3. T. Barrett, *et al.*, NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**, D991-995 (2013).
4. Z. Duren, X. Chen, R. Jiang, Y. Wang, W. H. Wong, Modeling gene regulation from paired expression and chromatin accessibility data. *Proc Natl Acad Sci U S A* **114**, E4914–E4923 (2017).
5. R. S. N. Fehrmann, *et al.*, Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115–125 (2015).
6. P. Mertins, *et al.*, Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
7. B. Zhang, *et al.*, Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
8. J. E. Lewis, T. E. Forshaw, D. A. Boothman, C. M. Furdui, M. L. Kemp, Personalized Genome-Scale Metabolic Models Identify Targets of Redox Metabolism in Radiation-Resistant Tumors. *Cell Syst* **12**, 68-81.e11 (2021).
9. A. Montagud, *et al.*, Patient-specific Boolean models of signalling networks guide personalised treatments. *eLife* **11**, e72626.
10. E. J. Pereira, *et al.*, Sporadic activation of an oxidative stress-dependent NRF2-p53 signaling network in breast epithelial spheroids and premalignancies. *Sci Signal* **13**, eaba4200 (2020).
11. N. Pappireddi, L. Martin, M. Wühr, A Review on Quantitative Multiplexed Proteomics. *Chembiochem* **20**, 1210–1224 (2019).
12. M. J. Ellis, *et al.*, Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov* **3**, 1108–1112 (2013).
13. A. Thompson, *et al.*, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895–1904 (2003).
14. L. C. Gillet, *et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111.016717 (2012).

15. Y. Perez-Riverol, *et al.*, The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442–D450 (2019).
16. M. J. Alvarez, *et al.*, Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838–847 (2016).
17. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* **21**, 630–644 (2020).
18. N. Fortelny, C. M. Overall, P. Pavlidis, G. V. C. Freue, Can we predict protein from mRNA levels? *Nature* **547**, E19–E20 (2017).
19. S. Tasaki, *et al.*, Inferring protein expression changes from mRNA in Alzheimer's dementia using deep neural networks. *Nat Commun* **13**, 655 (2022).
20. M. Wilhelm, *et al.*, Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
21. J. C. Taggart, H. Zaubert, M. Selbach, G.-W. Li, E. McShane, Keeping the Proportions of Protein Complex Components in Check. *Cell Syst* **10**, 125–132 (2020).
22. M. Giurgiu, *et al.*, CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559–D563 (2019).
23. A. L. Richards, M. Eckhardt, N. J. Krogan, Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol Syst Biol* **17**, e8792 (2021).
24. C. D. Griffiths, A. J. Sweatt, K. A. Janes, Simulating coxsackievirus B3 infection with an accessible computational model of its complete kinetics. *STAR Protoc* **2**, 100940 (2021).
25. A. B. Lopacinski, *et al.*, Modeling the complete kinetics of coxsackievirus B3 reveals human determinants of host-cell feedback. *Cell Syst* **12**, 304–323.e13 (2021).
26. G. Ciriello, *et al.*, Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506–519 (2015).
27. D. P. Nusinow, *et al.*, Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387–402.e16 (2020).
28. Y. Liu, *et al.*, Impact of Alternative Splicing on the Human Proteome. *Cell Rep* **20**, 1229–1241 (2017).
29. G. Rosenberger, *et al.*, A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data* **1**, 140031 (2014).
30. H. Srivastava, *et al.*, Protein prediction models support widespread post-transcriptional regulation of protein abundance by interacting partners. *PLoS Comput Biol* **18**, e1010702 (2022).
31. M. Ghandi, *et al.*, Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).

32. D.-G. Mun, *et al.*, Proteogenomic Characterization of Human Early-Onset Gastric Cancer. *Cancer Cell* **35**, 111-124.e10 (2019).
33. A. Ruepp, *et al.*, CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res* **38**, D497–D501 (2010).
34. E.-J. Wagenmakers, S. Farrell, AIC model selection using Akaike weights. *Psychon Bull Rev* **11**, 192–196 (2004).
35. M. Wühr, *et al.*, Deep Proteomics of the *Xenopus laevis* Egg using an mRNA-derived Reference Database. *Curr Biol* **24**, 1467–1475 (2014).
36. D. Szklarczyk, *et al.*, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638–D646 (2023).
37. J. Zecha, *et al.*, Peptide Level Turnover Measurements Enable the Study of Proteoform Dynamics. *Mol Cell Proteomics* **17**, 974–992 (2018).
38. Y. Liu, *et al.*, Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol* **37**, 314–322 (2019).
39. A.-D. Brunner, *et al.*, Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* **18**, e10798 (2022).
40. J. Reimegård, *et al.*, A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun Biol* **4**, 624 (2021).
41. M. Wang, C. J. Herrmann, M. Simonovic, D. Szklarczyk, C. von Mering, Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).
42. B. Eraslan, *et al.*, Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol Syst Biol* **15**, e8513 (2019).
43. T. Guo, *et al.*, Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. *iScience* **21**, 664–680 (2019).
44. W. C. Reinhold, *et al.*, RNA sequencing of the NCI-60: Integration into CellMiner and CellMiner CDB. *Cancer Res* **79**, 3514–3524 (2019).
45. K. Champi, *et al.*, Convergent network effects along the axis of gene expression during prostate cancer progression. *Genome Biol* **21**, 302 (2020).
46. S. Y. Ow, *et al.*, iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly.” *J Proteome Res* **8**, 5347–5355 (2009).
47. K. A. Janes, An analysis of critical factors for quantitative immunoblotting. *Sci Signal* **8**, rs2 (2015).
48. K. S. Kim, G. Hufnagel, N. M. Chapman, S. Tracy, The group B coxsackieviruses and myocarditis. *Rev Med Virol* **11**, 355–368 (2001).

49. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
50. C.-F. Liu, *et al.*, Whole-Transcriptome Profiling of Human Heart Tissues Reveals the Potential Novel Players and Regulatory Networks in Different Cardiomyopathy Subtypes of Heart Failure. *Circ Genom Precis Med* **14**, e003142 (2021).
51. M. Heinig, *et al.*, Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol* **18**, 170 (2017).
52. H. Fechner, *et al.*, Induction of coxsackievirus-adenovirus-receptor expression during myocardial tissue formation and remodeling: identification of a cell-to-cell contact-dependent regulatory mechanism. *Circulation* **107**, 876–882 (2003).
53. T. Kaur, *et al.*, Expression of coxsackievirus and adenovirus receptor and its cellular localization in myocardial tissues of dilated cardiomyopathy. *Exp Clin Cardiol* **17**, 183–186 (2012).
54. T. H. Dunnebacke, M. B. Reaume, Correlation of the yield of poliovirus with the size of isolated tissue cultured cells. *Virology* **6**, 8–13 (1958).
55. K. A. Hoadley, *et al.*, Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
56. R. A. Neff, *et al.*, Molecular subtyping of Alzheimer’s disease using RNA sequencing data reveals novel mechanisms and targets. *Sci Adv* **7**, eabb5398 (2021).
57. R. O. Ramirez Flores, *et al.*, Consensus Transcriptional Landscape of Human End-Stage Heart Failure. *J Am Heart Assoc* **10**, e019667 (2021).
58. J. S. Parker, *et al.*, Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol* **27**, 1160–1167 (2009).
59. C. M. Perou, *et al.*, Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
60. Cancer Genome Atlas Network, Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
61. R. S. Finn, *et al.*, Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* **375**, 1925–1936 (2016).
62. D. J. Slamon, *et al.*, Overall Survival with Ribociclib plus Fulvestrant in Advanced Breast Cancer. *N Engl J Med* **382**, 514–524 (2020).
63. R. S. Finn, *et al.*, Biomarker Analyses of Response to Cyclin-Dependent Kinase 4/6 Inhibition and Endocrine Therapy in Women with Treatment-Naïve Metastatic Breast Cancer. *Clin Cancer Res* **26**, 110–121 (2020).

64. N. C. Turner, *et al.*, Cyclin E1 Expression and Palbociclib Efficacy in Previously Treated Hormone Receptor-Positive Metastatic Breast Cancer. *J Clin Oncol* **37**, 1169–1178 (2019).
65. C. R. John, *et al.*, M3C: Monte Carlo reference-based consensus clustering. *Sci Rep* **10**, 1816 (2020).
66. B. Schwanhäusser, *et al.*, Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
67. C. Vogel, *et al.*, Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* **6**, 400 (2010).
68. R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, J. B. Hogenesch, A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci U S A* **111**, 16219–16224 (2014).
69. J. Lever, M. Krzywinski, N. Altman, Regularization. *Nature Methods* **13**, 803–804 (2016).
70. H. Haken, Information compression in biological systems. *Biol Cybern* **56**, 11–17 (1987).
71. E. Schneidman, M. J. Berry, R. Segev, W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
72. E. Brunk, *et al.*, Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272–281 (2018).
73. R. Agren, *et al.*, Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* **8**, e1002518 (2012).
74. S. A. Becker, B. O. Palsson, Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4**, e1000082 (2008).
75. S. Opdam, *et al.*, A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst* **4**, 318–329.e6 (2017).
76. C.-Q. Zhong, *et al.*, Generation of a murine SWATH-MS spectral library to quantify more than 11,000 proteins. *Sci Data* **7**, 104 (2020).
77. B. Efron, R. Tibshirani, Statistical data analysis in the computer age. *Science* **253**, 390–395 (1991).
78. R. Gao, *et al.*, Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39**, 599–608 (2021).
79. G. La Manno, *et al.*, RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
80. W. Saelens, R. Cannoodt, H. Todorov, Y. Saeys, A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547–554 (2019).

81. M. C. Chambers, *et al.*, A Cross-platform Toolkit for Mass Spectrometry and Proteomics. *Nat Biotechnol* **30**, 918–920 (2012).
82. H. L. Röst, *et al.*, OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741–748 (2016).
83. J. Teleanu, *et al.*, DIANA--algorithmic improvements for analysis of data-independent acquisition MS data. *Bioinformatics* **31**, 555–562 (2015).
84. H. L. Röst, *et al.*, TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat Methods* **13**, 777–783 (2016).
85. R. Leinonen, H. Sugawara, M. Shumway, The Sequence Read Archive. *Nucleic Acids Res* **39**, D19–D21 (2011).
86. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
87. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. S. Kovaka, *et al.*, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, 278 (2019).
89. G. H. Putri, S. Anders, P. T. Pyl, J. E. Pimanda, F. Zanini, Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
90. J. Morales, *et al.*, A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* **604**, 310–315 (2022).
91. D. P. Nusinow, S. P. Gygi, A Guide to the Quantitative Proteomic Profiles of the Cancer Cell Line Encyclopedia. 2020.02.03.932384 (2020).
92. Gene Ontology Consortium, *et al.*, The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
93. W. Shao, *et al.*, Comparative analysis of mRNA and protein degradation in prostate tissues indicates high stability of proteins. *Nat Commun* **10**, 2524 (2019).

Chapter 5:

Discussion

Acknowledgements

Much of the work presented in this chapter could not have been done without the help of Cameron D. Griffiths, PhD, my virus buddy and a scientific role model; Brent French, PhD, my committee chair and AAV9 extraordinaire; Page Murray and Mia Pergola, two wonderful UVA undergrads that I had the pleasure to mentor during their 3rd and 4th year; and Mohammad Aga, a summer REU student who honestly crushed it.

5.1 Summary, limitations, broader impacts, and future directions

Viruses are important pathogens. Disease results from complex interactions among the virus, host cell, and immune cells. Though we know much about each component, the link between viral infection and disease is still elusive (1). In this dissertation, I used systems-biology approaches to study virus–host interactions with one virus that is perhaps best-poised for such approaches, coxsackievirus B3 (CVB3). Below, I summarize the two major research projects and locally discuss their limitations, broader impacts, and future directions. I then discuss three projects for which data were generated and potential future follow-up before concluding this dissertation.

An accessible complete kinetic model for CVB3

In chapter 2, I described a complete kinetic model for CVB3 infection that encodes viral entry, replication, and encapsidation overlaid with innate-immune effector negative feedbacks (Figure 2.1A). The model was constructed with minimal parameter fitting, with 90+% of parameters derived from the literature or our own experiments, and each module was validated experimentally (Figures 2.3–2.9). With the validated model, we predicted a nonlinear association between i) the timing of interferon (IFN) signaling and ii) host-cell resistance to viral-proteinase antagonism of the host-cell response. We predicted that cells with baseline resistance to CVB3 proteinases have decreased sensitivity to IFN at late time points, whereas cells with increased resistance have sustained sensitivity (Figures 2.11A and 2.11B). To experimentally model host-cell resistance, we mutated MAVS at position 271 (Q271A) to confer resistance to cleavage by the 3C^{pro} and engineered transformed cardiomyocytes to express MAVS wildtype or Q271A upon doxycycline induction (Figures 2.11C–F). When cells were induced, infected, and treated with IFN, we saw the same pattern of time-dependent sensitivity (Figure 2.11I), indicating a role for MAVS as a critical determinant of the enteroviral response in human cells.

The results prompted a closer look at MAVS. The Q271A mutation was artificial, but we discovered a Q93E mutation that occurs in ~50% of people of East Asian ancestry (2). The mutation was significant because Q⁹³ was predicted to be susceptible to cleavage by 3C^{pro} while E⁹³ was not (3). We thus hypothesized that cells expressing MAVS Q⁹³ would have increased susceptibility to infection due to double the number of cleavage sites (Q⁹³ + Q²⁷¹). Strikingly, the opposite was true—cells expressing MAVS E⁹³/Q²⁷¹ had increased viral load (Figure 2.13D). The answer to this conundrum lied in the kinetics of MAVS polymerization—signaling, cleavage, and mitophagy. Singly-cleavable MAVS (E⁹³/Q²⁷¹) switched to a first-order decay process later than doubly-cleavable MAVS (Q⁹³/Q²⁷¹) but from a lower starting point, ultimately leading to less sustained interferon-stimulated gene

(ISG) production (Figure 2.13F) to fight the infection. Thus, the singly-cleavable allele would have increased viral load, as observed experimentally. Altogether, the results indicated that small biochemical differences are important for understanding complex host–virus interactions.

In constructing the mechanistic model, we made major simplifications in i) the encapsidation module, ii) the formation of viral replication organelles (VROs), and iii) the host-cell immune response, providing important limitations on the model.

Encapsidation is the least understood process of the CVB3 life cycle, and details of virion morphogenesis are unclear for enteroviruses broadly (4). One major question surrounds the details about how +vRNA is packaged into the capsid. Evidence suggests that some enteroviruses (e.g., parechovirus 1, foot-and-mouth disease virus) contain a packaging signal within their RNA that is required for efficient encapsidation (5, 6) while others (e.g., poliovirus, coxsackievirus B3) do not but require interactions between 2C^{ATPase} and VP3 capsid protein (7, 8). The encapsidation module for CVB3 was encoded with the best evidence to date but should be revisited as mechanistic details become clearer.

As the capsid forms, the evolving geometry should affect the addition of subsequent (pentamer) subunits due to the increased number of binding sites and the avidity effect. Further, the addition need not necessarily be a growing-capsid with a single pentamer—a 3-mer plus a 4-mer would form a 7-mer and have a greater avidity than a 2-mer with a 2-mer. Tracking the combinatorics would be computationally intractable, so we abstracted the process and assumed addition of one pentamer at a time with an unchanging weak contact affinity (9). The simplification was sufficient for our purposes but may not be for other enteroviruses where details of morphogenesis are different. For example, human rhinovirus (HRV) displays an inverse temperature dependence, replicating better at 33°C than 37°C (10–12), and more-detailed models of encapsidation may be required to capture the thermodynamic dependence. Assumptions of the encapsidation module should be refined as the data and computational tools become available and should be given extra attention when expanding to other enteroviruses.

VROs have been studied for decades and observed for many positive-strand RNA viruses (13–18), yet details of their nucleation are unclear due to technical limitations (19). Since data about VRO formation are lacking, we encoded a transition from cytoplasmic-volume to VRO-surface biochemistry when a threshold of 25 3D^{pol} molecules was met. (See Chapter 2.3.3 for details and explanation.) The approach was biologically artificial. Rather, one could encode the co-option of host-cell membranes by viral proteins 2BC and 3A (20, 21), where membranes transition from a “host-cell” compartment to a “VRO” compartment dictated by the concentration of 2BC–3A:

$[membrane_{VRO}] = [membrane_{host-cell}] \times [2BC - 3A]$. VRO formation could be modeled by early formalisms of nucleation–elongation (22), and viral replication machinery would be placed on the expanding VRO surface. VRO formation would be bound by experimental evidence of VRO surface area present by 4+ hours post infection (19), providing restrictions on the speed with which nucleation–elongation must occur.

Of course, a mechanistic encoding of VRO formation in this way could be too simple to be useful (23). Sometimes, expanding a model to include details is required to recapitulate observed systems-level properties. While it is unclear which details might be required for VRO formation, one place to start would be viral protein-mediated recruitment of the host-cell proteins Arf1 and GBP1. Arf1 and GBP1 are necessary for the co-option of phosphatidylinositol-4-phosphate (PI4P) lipids that build enteroviral VROs (24). Since the VRO transition is critical for enteroviral replication, such mechanistic explanations would be of broad importance for virologists.

Lastly, the host-cell immune response was greatly simplified by lumping the entire response into three negative feedback mechanisms (Figure 2.1A). Hundreds of ISGs are expressed in response to type I IFN signaling (25), but the antiviral proteins can be classified at a high level into a handful of main effector functions that we encode (26). However, MDA-5 is also an ISG (27) and signals through MAVS (28), establishing a positive feedback loop that drives the cell into an antiviral state. In the CVB3 model, we neglected positive feedback loops. Further, we limit negative feedback of the host-cell response to antagonism by viral proteinases, but reactive oxygen species produced during the antiviral response can degrade MAVS (29), providing a host-cell negative feedback mechanism on innate-immune signaling. Modeling every piece of the host-cell immune response would be impossible and not useful for our complete kinetic model (23); however, a simple expansion of a single positive feedback and negative feedback could be quickly implemented to include the host-cell contribution to regulation of the antiviral response.

The work had a major experimental caveat, as well. The MAVS-overexpression cell line retained endogenous MAVS, and induced expression was at a very high level (Figure 2.11C and 2.11D). It is unclear if the results would be the same if MAVS was expressed on a MAVS^{-/-} background and at endogenous levels. In future iterations, it would be good to perform a CRISPR/Cas9 knockout of MAVS and add back the alleles at near-endogenous levels.

The enterovirology community benefits greatly from the modular organization of the complete kinetics model because it can be quickly adapted to other enteroviruses. Within the genus, the most substantive differences lie in binding and entry. As an example, the CVB3 delivery module was swapped for poliovirus and infection simulated for 24 hours with or without additional IFN at 1.5 hours post-infection (Figure 5.1). Unlike CVB3, poliovirus binds to one receptor (PVR) and enters the cytoplasm before the virus–receptor complex has fully internalized (21, 30, 31). As expected, the faster delivery mechanism had a qualitative impact on viral progression with added interferon. In CVB3, interferon added 1.5 hours post-infection abrogated viral progression (Figure 5.1A). Conversely, interferon at 1.5 hours post-infection in poliovirus was not early enough to overcome the faster kinetics, and the end-state virion production was the same with and without interferon (Figure 5.1B). Other picornaviruses are more complicated, such as rhinoviruses that cause the common cold. Not only do rhinoviruses use different entry receptors, they display an inverse temperature dependence, replicating better at 33°C than 37°C (10–12). By adapting the complete kinetics model to human rhinovirus (via delivery module swapping) and incorporating a temperature-dependent component to the model, we could pinpoint the piece of the viral life cycle that benefits from slower kinetics (preliminary data discussed below). If experimentally validated, such a finding would provide one of the largest conceptual leaps in virology.

The “swappability” is also useful for studying emerging enteroviruses (32). For example, echovirus 11 has had (sometimes fatal) outbreaks historically (33–35) and recently (36, 37), yet there are few studies into its disease mechanism (38, 39). Such viruses could be modeled by complete kinetics to help experimentalists triangulate critical pieces of the viral life cycle that may be responsible for disease.

Generally speaking, the virology community is not comfortable (yet!) with computational modeling, though there is great appreciation for the approach. Often, modern virologists are not computationalists, and therefore models are inaccessible. The burden is on computationalists to provide tools that bring modeling to the virology community. To that end, we recast the complete kinetics model as a graphical user interface (GUI; see Chapter 3 for details). We prioritized the user’s ability to make changes to key parameters that differ between enteroviruses and host-cell systems so that the

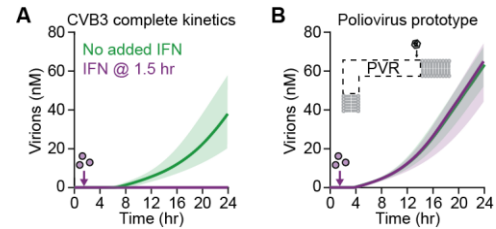


Figure 5.1. The complete kinetics model can be quickly adapted to other viruses. (A) A complete kinetics simulation with interferon (IFN) added 1.5 hours post-infection (purple) or without (green). (B) The same simulation as (A) but with parameters and equations for poliovirus binding–entry using its receptor, PVR. Predictions are shown as the median simulation \pm 90% nonparametric confidence interval from 100 runs with a parameter coefficient of variation of 5%.

model could be rapidly altered to study other enteroviruses. By converting to a GUI, we hope to encourage noncomputationalists to use the model for their own research and education purposes.

Protein copy-number estimation from transcriptomics

In chapter 4, I described a method for obtaining protein copy-number estimates from RNA-seq data by building statistical models that relate protein abundance to transcript abundance in paired data for >4000 genes (Figure 4.1). I found that protein abundance was described by one of three models: i) no dependence on its transcript (i.e., a constant value), ii) a nonlinear dependence on its transcript abundance, or iii) a dependence on its transcript abundance + the abundance of other genes (Figures 4.1C–E and 4.2C–F). The suite of models provided a platform to make proteome-wide copy number estimation from input RNA-seq data, a method we call protein inference from RNA (Pinferna). We validated Pinferna predictions by using independent data from dozens of cell lines and primary tissues (Figure 4.5). The validated suite of mRNA-to-protein relationships was considered complete and packaged as a web tool (<http://janeslab.shinyapps.io/Pinferna>).

We estimated the abundance of the CVB3 entry receptors CD55 and CXADR in 1489 human heart samples. When the complete kinetics of each were simulated, we uncovered scenarios of complete resistance to infection and three modes of productive infection (Figure 4.7C). As a comparison, we linearly scaled the mRNA abundances about a quantity of CD55 and CXADR randomly selected from the training data and categorized predictions as resistant, sublytic, or lytic. Some cases matched the true inferences, but most randomizations tended to predict ~100% resistance or ~100% lytic infection and underestimate the sublytic group (Figure 4.8H). We further checked whether CD55, CXADR, or both were the main contributor to susceptibility. By transcriptomics, *CD55* is much more widely distributed than *CXADR* (Figure 4.7D and 4.7E), hinting it may be the dominant receptor. However, by protein, CXADR has a much larger distribution and was the only receptor to stratify susceptibility (Figure 4.7F and 4.7G). The results indicated that mRNA inferences can misconstrue outputs of models built for protein networks.

As a second example application of Pinferna, we asked how the intrinsic molecular subtype of breast cancers would change when assigned by predicted proteomics. We obtained the transcriptomics of 796 breast cancers from The Cancer Genome Atlas, predicted their proteomics, and clustered them by both transcriptomics and predicted proteomics. Most sample assignments were unchanged (Figure 4.10A). Of those reassigned, we noted an enrichment for luminal subtypes (Figure 4.10B). We surveyed the

genes that drove reassignments and noted that they tended to have interactions with other genes that strongly influenced their predicted abundance (Figures 4.10C–F). Of particular interest was cyclin-dependent kinase (CDK) 4 because CDK inhibitors are approved to treat luminal breast cancers, but responsiveness has not consistently associated with the abundance of *CDK4* mRNA (40, 41). The results indicate a role for combining transcriptomics and predictions of the proteome to study cancer subtypes.

Pinferna is a method and a tool with far-reaching implications for systems biologists who often require protein abundances but must substitute RNA-seq (42–44). There are methods for obtaining protein copy numbers, such as quantitative immunoblotting (45) and ultra-sensitive selected reaction monitoring mass spectrometry (SRM) (46). Immunoblotting, however, requires antibodies and recombinant proteins for the quantitative standard. The conditions may be met for a handful of proteins, but scalability becomes an issue when dozens of protein abundances are required. SRM, and other mass spectrometry methods, are high-throughput, but they are technically challenging and often inaccessible. Mass spectrometry is not yet a commoditization like RNA-seq. Thus, for modelers concerned with protein estimates, inference methods such as Pinferna provide a scalable alternative to limited experimental techniques.

Pinferna outcompeted alternative methods for estimating protein abundances (Figures 4.5C–E), but the method is not without limitations. First, we made the assumption that cellular protein abundance is lognormally distributed about 10,000 copies per cell. The assumption was the same as that of the developer of SWATH (47), so we believe it is a valid approximation for an average cell. However, if protein abundances in a larger cell were distributed about 50,000 copies (48), the predictions would likely be off by fivefold. Future iterations of Pinferna could update the calibration datasets to include cell size as a factor in selecting the distribution of protein abundances and use cell size as input to appropriately scale the estimations.

Second, the validation data are SWATH measurements from public repositories and thus contingent upon all the assumptions of SWATH. A more rigorous validation would be against a select set of proteins quantified by immunoblotting. The number of genes in each category (M, HL, HL+LASSO; Figure 4.1C–E) is large enough to select proteins that meet the conditions for quantitative immunoblotting. Of course, the best outcome is that the results corroborate the predictions for all proteins. However, we may find that a certain class is predicted better than others, or discover that certain cell lines can be better predicted than others. Either way, the results would provide valuable information useful for updating Pinferna.

Finally, Pinferna is limited in its number of models due to relying on SWATH for calibration. SWATH is currently the best method for absolute proteomics, but relative methods such as tandem mass tag (TMT) mass spectrometry can quantify 8,000+ proteins (49). More proteins could be included by increasing the number of SWATH calibration datasets and taking the union of all measured proteins, but generating (or finding) SWATH datasets is challenging. Another potential way to increase the number of relationships would be to estimate a model for (currently “model-less”) genes based on the model of similar genes. For example, genes involved in cytoplasmic translation were highly enriched as M genes with no dependence on transcript abundance (Figure 4.3A). It is reasonable to assume, therefore, that other cytoplasmic-translation genes will not have a dependence on their mRNA levels. Deciding the protein abundance presents a larger challenge. One possibility would be to aggregate the abundances of all cytoplasmic-translation genes and choose the median \pm bootstrapped 95% confidence interval. For assumed HL and HL+LASSO genes, the parameter-estimation approach is more challenging. One could assess whether genes belonging to the same ontology have similar parameter sets that are distinct from genes other ontologies. If so, the estimated parameter set for a gene would be the median parameter set of similar genes. Though there are experimental and potentially computational avenues to increasing the suite of mRNA-to-protein relationships, the current number is still remarkable given experimental constraints. Altogether, the limitations of Pinferna do not preclude its use or overshadow its contribution to the field as a scalable alternative to estimating protein copy numbers.

5.2 Unpublished, but never forgotten

In vivo transduction of MAVS alleles to study their physiological impacts

Introduction

I previously demonstrated that MAVS is a crucial node for CVB3 infection dynamics (Figures 2.11 and 2.13) (50). Human MAVS (hMAVS) is cleaved at two sites—Gln⁹³ and Gln²⁷¹—that separate the CARD from the phosphorylation and mitochondrial-docking domains (Figure 2.13A) (50). The Gln²⁷¹ site arises from a six amino acid insertion resulting from a splice-site mutation that is conserved in humans and old-world monkeys (51). New-world monkeys lack the insertion and are also seronegative for CVB3 (52). By contrast, Gln⁹³ is widely conserved in mammals except in humans of East Asian ancestry, who have ~50% prevalence of the Glu⁹³ minor allele (rs17857295) shown to increase viral load in cells (2, 50). Interestingly, East Asia routinely struggles with picornaviral outbreaks (53), perhaps as a result of the minor allele prevalence. The observations warrant study into how the different allelotypes modulate CVB3-associated disease. *In vitro*, the Gln⁹³ allele

causes a ~40% reduction in viral load after 24 hours of infection (Figure 2.13C) (50). Over the course of one week, a daily serial 40% reduction would correspond to a $(1-0.40)^7 \cong 36$ -fold change in viral burden. Likely, the real fold change is larger because the life cycle for CVB3 is ~8 hours. Thus, even though viral burden seems to be only moderately reduced by the altered cleavage site, the accumulated effects during the course of viral infection may have drastic impacts on disease progression *in vivo*.

It is well known that CVB3 causes myocarditis in mice (54). It is also known that Mavs (mMavs, for the mouse allele) is required for an immune response in mice (55) and can modulate disease. It is not known how myocarditis, CVB3, and hMAVS/mMavs are coordinately linked. One study in mMavs knockout (mMavs^{-/-}) mice found that the mice were more susceptible to CVB3, but disease was monitored in the pancreas and liver, not the heart (56). In another *in vivo* study, different hMAVS allelotypes were delivered to the livers of mice and found to make a difference in combination with an interferon-based antiviral therapy (57). However, the latter study had three caveats that distinguish it from my study. First, the authors were studying hepatitis B virus—not a picornavirus. Second, they explored an allelotype different from the Glu⁹³/Gln⁹³ polymorphism that I identified. Third, they used mice that retained mMavs, which was sufficient for their purposes, but would be insufficient for my purposes. Still, their results demonstrated that hMAVS allelotypes coordinately alter the type I interferon response in tandem with stimulation from type I interferon—the same finding as my study with hMAVS (Figure 2.11A, 2.11B, and 2.11I). Thus, *in vivo* studies that link hMAVS allelotypes to CVB3-induced myocarditis are unavailable.

Viral myocarditis has long been known to depend on both viral infection and the adaptive immune response (58). Studies in immunodeficient mice infected with CVB3 show a clear cardiopathogenesis for the virus alone (59–61). The role of adaptive immunity in myocarditis is more complex. In immunocompetent mice, myocarditis tends to be worse following CVB3 infection in some cases and attenuated in others. Attenuated myocarditis is associated with cardiac infiltration of anti-inflammatory M2 macrophages and regulatory T cells (61–64). Macrophages and T cells are recruited by cytokines and chemokines, indicating the cytokine/chemokine landscape produced by infected cells plays a role in the progression of myocarditis. For example, the cytokines tumor necrosis factor-alpha (TNF- α) and interleukin-1 (IL-1) have been shown to modulate myocarditis severity in infections with both cardio- and non-cardiopathogenic viruses (62, 65). Together, the literature suggests that the cytokine/chemokine-coordinated recruitment of innate-immune and adaptive-immune infiltrates modulates myocarditis.

The link between MAVS and adaptive immunity is undefined, but clear, for CVB3. MAVS is responsible for relaying signals that cause production of cytokines and chemokines. The strength and duration of MAVS signaling could alter the cytokine/chemokine landscape, suggesting that regulation of MAVS can alter the recruitment of macrophages and adaptive-immune cells. Indeed, a study in *mMavs*^{-/-} mice demonstrated a weakened T cell response to a non-picornavirus (66), but no such link has been explored for CVB3. Further, much of the data gathered for adaptive immunity has come from mice that retained *mMavs*. Given the differences with hMAVS, there is a need to explore how hMAVS allelotypes alter the adaptive immune landscape and attenuate CVB3 to reduce myocarditis.

Results

We obtained *Mavs*^{-/-} pups and injected them retro-orbitally with 3×10^{10} genome copies of AAV2/9 containing V5-MAVS-P2A-3xFLAG-luciferase (luc) under control of the cTnT promoter to ensure heart-specific expression (67). After four weeks, stable expression was confirmed by bioluminescence imaging (BLI; Figure 5.2A). BLI measured luc expression, so we prepped heart tissue for immunoprecipitation and immunocytochemistry. By both methods, we confirmed sex-independent expression of both 3xFLAG-luc and V5-MAVS in an AAV2/9-transduced heart (Figure 5.2B–E).

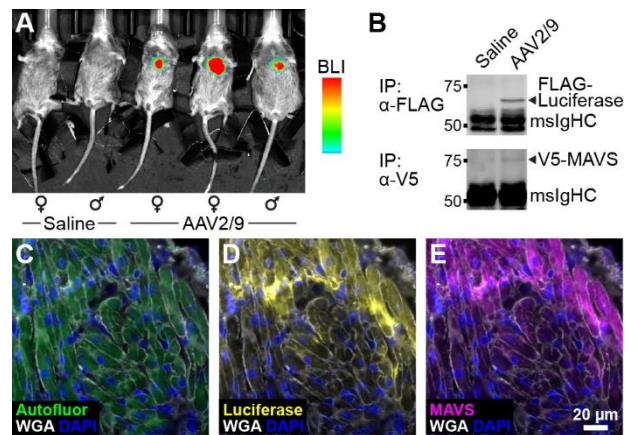


Figure 5.2. Proof of concept for AAV2/9 transduction of V5-MAVS-P2A-3xFLAG-luciferase (luc) into *Mavs*^{-/-} mice. (A) Bioluminescence imaging (BLI) of *Mavs*^{-/-} mice transduced with AAV2/9 or saline. (B) Immunoprecipitation of 3xFLAG-luc and V5-MAVS from *Mavs*^{-/-} hearts. (C–E) Immunocytochemistry for luc (yellow) and MAVS (magenta) in *Mavs*^{-/-} hearts. Autofluorescence (Autofluor, green) was used to visualize all cells, DAPI to label nuclei, and wheat germ agglutinin (WGA) to label membranes.

Acknowledgements: Dr. Brent French provided the parental AAV2/9 vector. Dr. Cameron Griffiths helped with all experiments to generate this data. Dr. Kevin Janes assembled this figure.

Discussion and future directions

The work here displays only pilot data during optimization of the protocol. The original goal was to use the same MAVS alleles as were used in cells (see chapter 2). In future experiments, mice will be infected with CVB3 after checking for stable MAVS expression via BLI then euthanized at peak disease 7–10 post infection. The prepped heart will be used for biochemical analysis (western blot, qPCR) of MAVS cleavage and ISG expression, virological analysis (plaque assay) to quantify viral load, and histological

analysis (immunohistochemistry) to assess myocarditis and immune cell infiltration. These experiments will demonstrate how allelotypic differences in hMAVS alter CVB3 progression and resulting disease by uniquely linking multiple levels of analysis.

Methods

An empty AAV2/9 vector containing 3xFLAG-luc was provided by Dr. Brent French at UVA. V5-MAVS-P2A was cloned into the vector by PCR, and the sequence-verified plasmid was shipped to Vector Biolabs for packaging into AAV2/9 capsids.

Mavs^{-/-} mice were obtained from Jackson labs (Stock No: 008634) and used to establish a breeding colony. Pups 7–10 days of age were anesthetized with isoflurane and injected retro-orbitally with 3×10^{10} genome copies of AAV2/9 in 10 μ L saline or with an equal volume of saline as a control. To check stable expression after 3–4 weeks, mice were injected intraperitoneally with luciferin and scanned using a Lago X scanner to observe bioluminescence.

Mice were euthanized in a CO₂ chamber, and the heart was quickly dissected to ensure integrity of the tissue. Heart tissue was cut into fifths (except for immunoprecipitation) and prepped by different methods: i) frozen in NEG-50 and isopentane, ii) flash-frozen in liquid nitrogen and ground into a powder for protein or RNA extraction, and iii) stored in All-Protect, a long-term storage buffer that stabilizes protein and RNA.

Fresh-frozen tissue was cryosectioned and sent to UVA's biorepository and tissue research facility (BTRF) for antigen retrieval via the low-pH method. The sections were then delivered to the Research Histology Core for H & E staining. Remaining sections were assessed for 3xFLAG-luc and V5-MAVS by immunocytochemistry.

The immunoprecipitation experiments required large quantities of starting material while the protocol was being optimized. Therefore, whole hearts were used in these final experiments. Whole hearts were flash-frozen in liquid nitrogen and ground with an ice-cold mortar and pestle on dry ice before being added to RIPA lysis buffer overnight at 4°C on the rotator. The next morning, the tissue was further ruptured by pipetting, incubated on ice for 20 min, then centrifuged at 16,800 RCF for 15 min. The supernatant was transferred to a fresh tube and the lysate quantified by BCA assay.

For the pulldown, 2 mg lysate was incubated with mouse-IgG as a control, ms-FLAG, or ms-V5. The Ips were blotted using the ms-anti-FLAG and ms-anti-V5 antibodies. The blot was stripped with 6 M guanidine-HCl at room temperature for 10 min, washed with PBS for 5 min, re-blocked, and re-probed for ck-anti-V5 (the more-sensitive V5 antibody).

A cold nose is better for the cold: Deciphering a mechanism for the inverse temperature dependence of human rhinovirus

Introduction

Human rhinoviruses (HRV) are the leading causative agents of the cold and a billions-of-dollars economic burden annually (68). Rhinoviruses are classified into two groups based on their cell-surface entry receptor—major group, which binds to ICAM-1 (69–71), and minor group, which binds to LDL-family receptors (LDLR) (72). Though HRV has been studied since the 1950s (68), a major question remains: Why does HRV replicate more efficiently at 33°C than 37°C (10–12)? From a teleological perspective, the preference for the colder temperature makes sense as HRV infects the cooler nasal epithelia. However, the observation is enigmatic mechanistically, as biochemical kinetics occur more slowly at lower temperatures. How, then, do slower kinetics equal greater replication? We proposed to use our mechanistic model to shed light on this inverse temperature-dependence conundrum.

Results

To begin, we chose HRV2 as a representative minor-group virus by suggestion of Dr. Monica Lawrence and confirmed inverse temperature-dependent replication (Figure 5.3A). To build the HRV2 model, we re-coded the entry module to reflect HRV2:LDLR binding and entry. The major challenge was to encode temperature dependence into the parameters.

Kinetic parameters were modulated from their base value using the Arrhenius equation:

$$k_{T_{new}} = k_{T_{base}} \exp \left[-\frac{E_A}{R} \left(\frac{1}{T_{new}} - \frac{1}{T_{base}} \right) \right],$$

where k is the kinetic parameter, T is temperature, E_A is activation energy, and R is the universal gas constant. The Arrhenius equation could not be used for the capsid

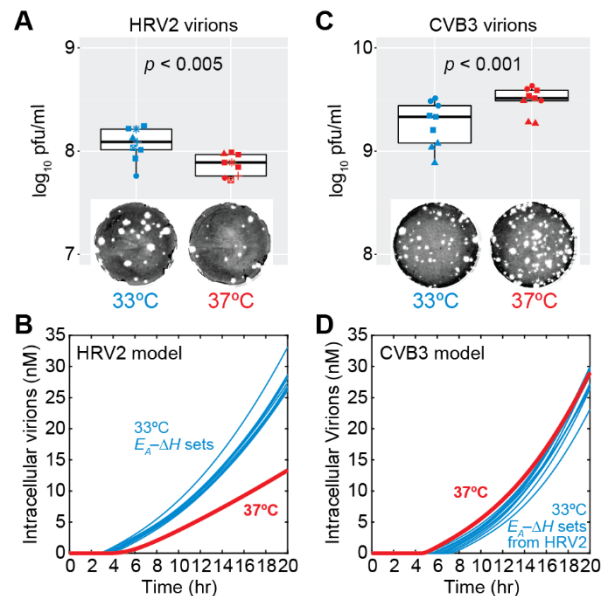


Figure 5.3. Inverse temperature dependence of human rhinovirus 2 (HRV2). (A) Plaque assays for HRV2 virions at 33°C (blue) or 37°C (red) (markers indicate different paired conditions). (B) Simulations of HRV2 complete kinetics using ten $E_A-\Delta H$ parameter sets that cause inverse temperature dependence. (C, D) Experimental and computational evidence that CVB3 does not exhibit inverse temperature dependence.

All experiments were performed in HeLa H1 cells, and the same $E_a-\Delta H$ parameter sets were used for both models.

Acknowledgements: Mia Pergola, Page Murray, and Dr. Cameron Griffiths helped encode temperature dependence into the model and reformulate the entry module for HRV2. Dr. Cameron Griffiths performed all experiments. Dr. Kevin Janes assembled this figure.

affinity parameters (in the encapsidation module) since they were not on–off kinetic parameters. Thus, we employed the van’t Hoff equation:

$K_{D,T_{new}} \approx K_{D,T_{base}} \exp \left[\frac{\Delta H}{R} \left(\frac{1}{T_{new}} - \frac{1}{T_{base}} \right) \right]$, where K_D is the equilibrium constant, ΔH is reaction enthalpy, T is temperature and R is the universal gas constant. For many parameters, E_A and ΔH were unknown. Rather than encode specific values, we sampled over a range of biologically plausible E_A values (2 to 40 kcal/mol) for kinetic parameters and ΔH values (–20 to +20 kcal/mol) for capsid affinities. For many E_A – ΔH sets, we were able to achieve inverse temperature dependence (Figure 5.3B). As a control, we performed simulations of CVB3 infection with the same parameter sets and remarkably found no inverse temperature dependence, consistent with experimental results (Figure 5.3C and 5.3D). We dove further into the parameter sets and noticed that ISG formation was recurrently among the most temperature sensitive (i.e., largest E_A) and, therefore, among the most reduced kinetic parameters. Thus, the results indicate that inverse temperature dependence arises from the faster entry of HRV2, allowing the virus to get ahead of the slowed-down ISG response.

Discussion and future directions

Another study has demonstrated that inverse temperature dependence arises for another minor-group rhinovirus due to slowed host-cell immune responses (12). Interestingly, they found that viral replication was enhanced when genetic deficiencies in MAVS slowed innate-immune signaling. The deficiency fits nicely with the results our kinetic model of MAVS signaling (Figure 2.13E and 2.13F). Genetic deficiencies that alter the kinetics of key innate-immune signalers thus represent another population-level difference that alter susceptibility to viral infection.

Future work will expand the model to include major-group rhinoviruses. Generally, the major-group viruses bind more slowly and less tightly than minor-group viruses (73, 74). Thus, the entry mechanics for major-group viruses may be slowed. Given the previous results that a fast entry gets the virus ahead of innate-immune signaling, it is unclear how the (potentially much) slower entry of major-group rhinoviruses will alter its replication efficiency. The inverse temperature dependence is still present for the major group, but perhaps the mechanism is different.

One major assumption that we made for the HRV model was the range of E_A and ΔH values. The model should be re-parameterized using experimentally determined values when possible. The “new” model would provide a platform to rapidly assess which component of the viral life cycle is responsible for the inverse temperature dependence,

providing a testable hypothesis. Identification of a mechanism would provide one of the largest conceptual leaps in virology.

To cleave or not to cleave: IRF3 is in question

Introduction

Previous work in MAVS demonstrated that altered propensity for cleavage by viral proteinases can have significant impacts on viral replication (Figure 2.11 and 2.13). I looked into other innate immunity regulators and found that interferon regulatory factor 3 (IRF3)—a transcription factor for ISGs that homodimerizes upon activation downstream of MAVS signaling (28, 75–78)—has an Arg⁹⁶ to Gln⁹⁶ polymorphism, with the minor allele yielding a predicted 3C^{pro} consensus site. Cleavage at Gln⁹⁶ would separate the N-terminal DNA-binding domain (DBD) from the C-terminal interferon-association domain responsible for dimerization (79). A Gln⁹⁶ cleavage could thus have a dominant-negative effect since both subunits of the homodimer need an active DBD (80), warranting a rigorous study into the effects of IRF3 cleavage similar to those in MAVS.

Results

I immunoprecipitated V5-tagged IRF3 Arg⁹⁶ or Gln⁹⁶, incubated the immunoprecipitates with recombinant 3C^{pro}, and found indications that Gln⁹⁶ has enhanced cleavage compared to Arg⁹⁶ (Figure 5.4A). However, the results were somewhat inconclusive as we had no measure for whether the recombinant 3C^{pro} was active and did not observe a cleavage

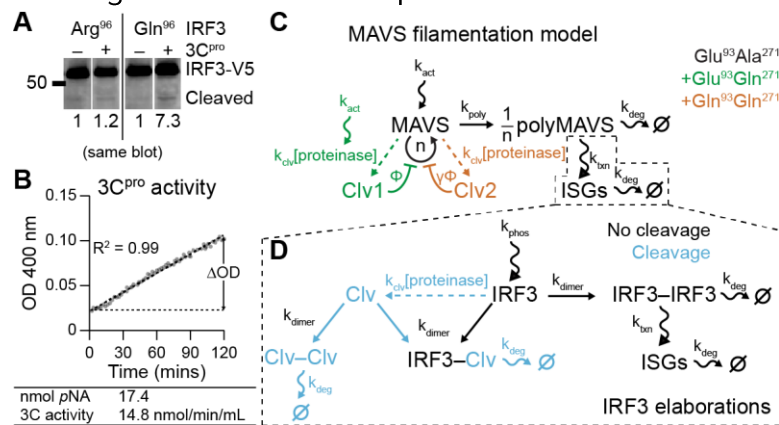


Figure 5.4. Implications of IRF3 Arg96 or Gln96 in innate immune signaling. (A) IRF3 Gln⁹⁶ has enhanced cleavage. V5-tagged IRF3 Arg⁹⁶ or Gln⁹⁶ was purified by IP pulldown and incubated with or without recombinant 3C^{pro}. Quantified densitometry is beneath with each allele normalized to each –3C^{pro} condition. (B) Recombinant 3C^{pro} is highly active. Cleavage of a 3C^{pro} substrate releases a chromophore (pNA) quantified at OD = 400 nm over time. Change in OD relates to activity of the proteinase. (C) Model for self-assembly of MAVS and cleavage by viral proteinase. MAVS is activated (k_{act}) and polymerizes (k_{poly}) into polyMAVS that signals (k_{txn}) to induce ISGs; both degrade at the same rate (k_{deg}). The base model (black) is elaborated with one (green) or two (brown) potential cleavage reactions. Viral proteinase is activated (k_{act}) and cleaves MAVS (k_{clv}) into cleavage products (Civ1 or Civ2) that inhibit MAVS polymerization with unequal feedback strength (ϕ and $\gamma\phi$). Boxed area is elaborated in (D). (D) Proposed IRF3 elaborations to the MAVS filamentation model (C). The base model (black) is elaborated with cleavage activity (blue). IRF3 is activated by polyMAVS (k_{phos}) and dimerizes (k_{dimer}) to form IRF3-IRF3 that transcribes ISGs (k_{txn}), both of which can degrade (k_{deg}). Viral proteinase cleaves IRF3 (k_{clv}) yielding cleaved IRF3 (Civ), which forms a homodimer with itself (Civ-Civ) or with noncleaved IRF3 (IRF3-Civ), both of which are inactive. Acknowledgements: Mohammad Aga performed the 3C^{pro} activity assay in (B).

fragment in the MAVS positive control (results not shown), prompting a more-rigorous analysis of 3C^{pro} activity.

We purified recombinant 3C^{pro} a second time and measured its activity using a chromogenic assay (Figure 5.4B; Methods). The proteinase was very potent, able to cleave 14.8 nmol of substrate/min/mL proteinase. We incubated saturating amounts of proteinase with purified IRF3 but again observed no clear indication of cleavage. We expressed both forms of IRF3 in cells and infected with CVB3, but a differential cleavage pattern was still inconclusive (data not shown). Early experiments indicate some preference for cleavage (Figure 5.4A), but subsequent attempts by IP and by infection in cells have different results. More attempts are needed to optimize the protocols and reagents used for the experiments.

Discussion and future directions

Differential cleavage of IRF3 could have potent effects on CVB3 replication due to the dominant negative effect it would have on the IRF3 dimer pool. However, it remains unclear whether CVB3 cleaves the minor allele of IRF3, despite the presence a consensus sequence for cleavage. Further attempts could be tried with some optimization.

In our experiments, V5-IRF3 was expressed in cells, purified by immunoprecipitation, and cleaved off of beads. In future attempts, IRF3 should be purified as was done for 3C^{pro}. Even though 3C^{pro} was added at (what was calculated to be) a saturating concentration, the recognition of the IRF3 cleavage site could be much slower than recognition of the substrate in the activity assay. Thus, we may have added too little proteinase or for too little of time to get robust cleavage assessable by immunoblotting. 3C^{pro} concentrations were maximal in the reaction volume, but longer incubation times are possible and may be necessary.

There is conflicting evidence in the literature as to whether 2A^{pro} or 3C^{pro} cleaves certain proteins. For example, both have been reported to cleave MAVS and result in similar cleavage patterns (81–83). Cleavage may be dependent on the experimental system (e.g., in a tube versus in cells). Though we assumed that 3C^{pro} is responsible for IRF3 cleavage, cleavage could be mediated by the 2A^{pro}. We have purified 2A^{pro}, but its activity needs to be assessed. Afterwards, the same experiment could be performed.

Another assessment of IRF3 cleavage could happen in cells. A previous undergraduate student (Page Murray) established transformed cardiomyocyte cell lines that stably express 3C^{pro}. A second transduction with IRF3 (Arg⁹⁶ or Gln⁹⁶) would result in a cell system that could be used to assess IRF3 cleavage. Similarly, a cell line expressing 2A^{pro} could be established, though previous attempts (also by Page) failed, presumably

due to toxicity of the proteinase though added in minimal amounts. A more-native environment for the proteinases and IRF3 may give rise to a differential cleavage pattern.

The lack of differential cleavage may be a technical issue. Thus, it would be beneficial to express the different forms of IRF3 in cells, infect with CVB3, and assess endpoint viral titers by plaque assay. The expectation is that cleavable IRF3 will have reduced innate-immune signaling, giving rise to increased viral titers. Further, the proteoforms of IRF3 could be co-expressed with the proteoforms of MAVS. The dual allele-specific susceptibility to cleavage could have interesting outcomes. I previously proposed an elaboration of the MAVS signaling model that would expand MAVS signaling to first activate IRF3 before production of ISGs (Figures 5.4C and 5.4D). The elaboration has not been done but could be rapidly implemented as a first-pass as to whether interest behavior might arise if the experiments were performed. For the MAVS–IRF3 combination model, I would expect IRF3 Gln⁹⁶ to dominate the response due to the dominant-negative effect.

Finally, I focused here on IRF3. However, a deeper bioinformatic search might reveal candidate cleavage sites that are created or destroyed in other innate-immune signalers. All are potential candidates for assessing how they impact the dynamics of CVB3 infection. The point ties in to the broader theme of population-level differences being used to explain differential outcome with disease.

Methods

HA-tagged 3C^{pro} was cloned in a pGEX-4T-1 vector for protein purification and transformed into bacteria. Proteins were purified as described in (84). For 3C^{pro}, we believed that the large GST protein would hinder its proteinase activity, so we cleaved 3C^{pro} off the column by incubating with thrombin.

The activity of 3C^{pro} was assessed with the HRV 3C Protease Activity Assay Kit from Abcam (ab211088) and following the provided protocol. The mass of pNA released was calculated by

$$nmol\ pNA = \frac{\Delta OD - b_{pNA\ standard}}{m_{pNA\ standard}}$$

where ΔOD is the difference in the absorbance value at two selected time points (Figure 5.4B), b is the intercept of the best-fit line through a pNA standard curve (not shown), and m is the slope of the best-fit line through the pNA standard curve (not shown). The mass of pNA was then used to calculate 3C^{pro} activity using the equation

$$3C\ activity = \frac{nmol\ pNA}{\Delta t \times V_{3C}}$$

where Δt is the difference in the chosen time points in min and V_{3C} is the volume of

recombinant 3C^{pro} added to the reaction in mL.

IRF3 Arg⁹⁶ or Gln⁹⁶ was cloned into pLX302-V5 and lipofected into 293T cells. Protein lysates were used in immunoprecipitation (IP) experiments for assessment of IRF3 cleavage by 3C^{pro}. To assess IRF3 cleavage, 500 µg lysate was incubated with V5 antibody overnight at 4°C then pulled down with beads. IRF3 was cleaved off the beads by incubating with 27 µL recombinant 3C^{pro} for 1 hr. Based on the activity of 3C^{pro}, this is enough volume to cleave 0.4 nmol substrate/min. The estimated amount of IRF3 was no more than 9.7 nmol, yielding a required time of 24 min to cleave everything. Thus, a 1-hr incubation time should completely saturate cleavage. Afterwards, everything was boiled off the beads and collected for immunoblotting.

5.3 Concluding remarks

“All models are wrong, but some are useful.”

–Kevin Janes (originally, I thought at the start of my PhD), systems biologists generally (I learned during my PhD), originally George Box (I learned when writing this dissertation)

In this dissertation, I proposed some useful models and discovered some interesting biology. I set out to identify population-level differences in host-cell proteins that are responsible for susceptibility to infection by CVB3. I identified two, MAVS and CAR.

The mechanistic model represents the first kinetic model (to my knowledge) of the complete life cycle of a virus. The work joins a long history of mathematical modeling of viral infection, and hopefully promotes interest in computational modeling among virologists. While the model simulated CVB3 specifically, it is emphasized that the model can be quickly adapted to other enteroviruses.

The findings with MAVS have implications for other proteins that may have cleavage-susceptible or cleavage-resistant alleles. Such genetic differences are useful for determining how small biochemical changes in virus–host interactions impact disease. One area of interest would be in assessing inborn errors of interferon-immunity (IEI), which have dysregulated type I/II interferon responses. Since we found that the timing of type I interferon signaling has substantial impacts on viral dynamics, people with IEIs represent a population that may have impaired innate-immune responses to enteroviruses.

It has been known for some time that CAR expression dictates susceptibility to CVB3. However, it was not known how CAR abundance is conditional on DAF abundance. In simulating infection in nearly 1500 human heart samples, we confirmed that population-level differences in CAR, but not DAF, drive susceptibility and stratify predicted severity of infection.

To obtain DAF–CAR values, I created Pinferna, a suite of 4366 statistical models able to estimate protein copy numbers from RNA-seq input. The work taught me that shifting among different types of models and blending their strengths builds a toolset that can be used to answer interesting questions—I couldn't have made the biological discovery regarding surface receptor-dependent CVB3 susceptibility without incorporating statistical modeling.

This work examined intracellular pathways. Going forward, I am broadly interested in incorporating intercellular networks in infectious disease. Disease is the culmination of pathogen–host-cell, host-cell–host-cell, and host-cell–immune-cell interactions. To begin building these networks, computational and experimental approaches will be needed... and someone willing to do the work. I am willing, and I believe the work presented in this dissertation has given me the necessary training to go forth and model.

5.4 References

1. Connecting the dots from viral infection to disease. *Nat Microbiol* **8**, 1363–1364 (2023).
2. L. Phan, *et al.*, ALFA: Allele Frequency Aggregator. *National Center for Biotechnology Information* (2020).
3. N. Blom, J. Hansen, D. Blaas, S. Brunak, Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci* **5**, 2203–2216 (1996).
4. P. Jiang, Y. Liu, H.-C. Ma, A. V. Paul, E. Wimmer, Picornavirus morphogenesis. *Microbiol Mol Biol Rev* **78**, 418–437 (2014).
5. G. Logan, *et al.*, Deep Sequencing of Foot-and-Mouth Disease Virus Reveals RNA Sequences Involved in Genome Packaging. *J Virol* **92**, e01159-17 (2017).
6. S. Shakeel, *et al.*, Genomic RNA folding mediates assembly of human parechovirus. *Nat Commun* **8**, 5 (2017).
7. C. Wang, P. Jiang, C. Sand, A. V. Paul, E. Wimmer, Alanine scanning of poliovirus 2CATPase reveals new genetic evidence that capsid protein/2CATPase interactions are essential for morphogenesis. *J Virol* **86**, 9964–9975 (2012).
8. Y. Liu, *et al.*, Direct interaction between two viral proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus morphogenesis. *PLoS Pathog* **6**, e1001066 (2010).
9. C. Li, J. C.-Y. Wang, M. W. Taylor, A. Zlotnick, In vitro assembly of an empty picornavirus capsid follows a dodecahedral path. *J Virol* **86**, 13062–13069 (2012).
10. E. J. Stott, G. F. Heath, Factors affecting the growth of Rhinovirus 2 in suspension cultures of L132 cells. *J Gen Virol* **6**, 15–24 (1970).

11. N. G. Papadopoulos, G. Sanderson, J. Hunter, S. L. Johnston, Rhinoviruses replicate effectively at lower airway temperatures. *J Med Virol* **58**, 100–104 (1999).
12. E. F. Foxman, *et al.*, Temperature-dependent innate defense against the common cold virus limits viral replication at warm temperature in mouse airway cells. *Proc Natl Acad Sci U S A* **112**, 827–832 (2015).
13. K. Bienz, D. Egger, L. Pasamontes, Association of polioviral proteins of the P2 genomic region with the viral replication complex and virus-induced membrane synthesis as visualized by electron microscopic immunocytochemistry and autoradiography. *Virology* **160**, 220–226 (1987).
14. K. Bienz, D. Egger, T. Pfister, M. Troxler, Structural and functional characterization of the poliovirus replication complex. *J Virol* **66**, 2740–2747 (1992).
15. M. Binder, *et al.*, Replication vesicles are load- and choke-points in the hepatitis C virus lifecycle. *PLoS Pathog* **9**, e1003561 (2013).
16. S. Dales, H. J. Eggers, I. Tamm, G. E. Palade, ELECTRON MICROSCOPIC STUDY OF THE FORMATION OF POLIOVIRUS. *Virology* **26**, 379–389 (1965).
17. S. Miller, J. Krijnse-Locker, Modification of intracellular membrane structures for virus replication. *Nat Rev Microbiol* **6**, 363–374 (2008).
18. C. Zitzmann, *et al.*, Mathematical modeling of plus-strand RNA virus replication to identify broad-spectrum antiviral treatment strategies. *PLoS Comput Biol* **19**, e1010423 (2023).
19. R. W. A. L. Limpens, *et al.*, The transformation of enterovirus replication structures: a three-dimensional study of single- and double-membrane compartments. *mBio* **2**, e00166-11 (2011).
20. D. A. Suhy, T. H. Giddings, K. Kirkegaard, Remodeling the endoplasmic reticulum by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-induced vesicles. *J Virol* **74**, 8953–8965 (2000).
21. V. Racaniello, “Picornaviridae: the viruses and their replication” in *Fields Virology*, 6th ed, D. M. Knipe, P. M. Howley, Eds. (Wolters Kluwer/Lippincott Williams & Wilkins Health, 2013).
22. C. Frieden, D. W. Goddette, Polymerization of actin and actin-like systems: evaluation of the time course of polymerization in relation to the mechanism. *Biochemistry* **22**, 5836–5843 (1983).
23. K. A. Janes, D. A. Lauffenburger, Models of signalling networks - what cell biologists can gain from them and give to them. *J Cell Sci* **126**, 1913–1921 (2013).
24. N.-Y. Hsu, *et al.*, Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* **141**, 799–811 (2010).

25. S. D. Der, A. Zhou, B. R. Williams, R. H. Silverman, Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A* **95**, 15623–15628 (1998).
26. A. J. Sadler, B. R. G. Williams, Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559–568 (2008).
27. D. Kang, *et al.*, mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A* **99**, 637–642 (2002).
28. F. Hou, *et al.*, MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* **146**, 448–461 (2011).
29. N. Qi, *et al.*, Multiple truncated isoforms of MAVS prevent its spontaneous aggregation in antiviral innate immune signalling. *Nat Commun* **8**, 15676 (2017).
30. B. Brandenburg, *et al.*, Imaging poliovirus entry in live cells. *PLoS Biol* **5**, e183 (2007).
31. B. M. McDermott, A. H. Rux, R. J. Eisenberg, G. H. Cohen, V. R. Racaniello, Two distinct binding affinities of poliovirus for its cellular receptor. *J Biol Chem* **275**, 23089–23096 (2000).
32. D. Lugo, P. Krogstad, Enteroviruses in the early 21st century: new manifestations and challenges. *Curr Opin Pediatr* **28**, 107–113 (2016).
33. C. Miwa, S. Sawatari, [Epidemic of echo 11 virus infection in Gifu Prefecture in 1993]. *Kansenshogaku Zasshi* **68**, 1251–1255 (1994).
34. J. R. Patel, J. Daniel, V. I. Mathan, An epidemic of acute diarrhoea in rural southern India associated with echovirus type 11 infection. *J Hyg (Lond)* **95**, 483–492 (1985).
35. S. Berkovich, S. Kibrick, ECHO II OUTBREAK IN NEWBORN INFANTS AND MOTHERS. *Pediatrics* **33**, 534–540 (1964).
36. J. Lu, *et al.*, Tracking echovirus eleven outbreaks in Guangdong, China: a metatranscriptomic, phylogenetic, and epidemiological study. *Virus Evol* **6**, veaa029 (2020).
37. Y.-C. Chen, *et al.*, Clinical characteristics of echovirus 11 and coxsackievirus B5 infections in Taiwanese children requiring hospitalization. *J Microbiol Immunol Infect* **54**, 581–587 (2021).
38. C. Wang, *et al.*, Echovirus 11 infection induces pyroptotic cell death by facilitating NLRP3 inflammasome activation. *PLoS Pathog* **18**, e1010787 (2022).
39. G. Zhang, *et al.*, Pathological Features of Echovirus-11-Associated Brain Damage in Mice Based on RNA-Seq Analysis. *Viruses* **13**, 2477 (2021).

40. R. S. Finn, *et al.*, Biomarker Analyses of Response to Cyclin-Dependent Kinase 4/6 Inhibition and Endocrine Therapy in Women with Treatment-Naïve Metastatic Breast Cancer. *Clin Cancer Res* **26**, 110–121 (2020).
41. N. C. Turner, *et al.*, Cyclin E1 Expression and Palbociclib Efficacy in Previously Treated Hormone Receptor-Positive Metastatic Breast Cancer. *J Clin Oncol* **37**, 1169–1178 (2019).
42. J. E. Lewis, T. E. Forshaw, D. A. Boothman, C. M. Furdai, M. L. Kemp, Personalized Genome-Scale Metabolic Models Identify Targets of Redox Metabolism in Radiation-Resistant Tumors. *Cell Syst* **12**, 68-81.e11 (2021).
43. A. Montagud, *et al.*, Patient-specific Boolean models of signalling networks guide personalised treatments. *eLife* **11**, e72626.
44. E. J. Pereira, *et al.*, Sporadic activation of an oxidative stress-dependent NRF2-p53 signaling network in breast epithelial spheroids and premalignancies. *Sci Signal* **13**, eaba4200 (2020).
45. K. A. Janes, An analysis of critical factors for quantitative immunoblotting. *Sci Signal* **8**, rs2 (2015).
46. P. Picotti, R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9**, 555–566 (2012).
47. Y. Liu, *et al.*, Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol* **37**, 314–322 (2019).
48. B. Schwanhäusser, *et al.*, Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
49. D. P. Nusinow, *et al.*, Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402.e16 (2020).
50. A. B. Lopacinski, *et al.*, Modeling the complete kinetics of coxsackievirus B3 reveals human determinants of host-cell feedback. *Cell Syst* (2021) <https://doi.org/10.1016/j.cels.2021.02.004>.
51. M. R. Patel, Y.-M. Loo, S. M. Horner, M. Gale, H. S. Malik, Convergent evolution of escape from hepatic viral antagonism in primates. *PLoS Biol* **10**, e1001282 (2012).
52. F. Deinhardt, A. Holmes, J. Devine, J. Deinhardt, Marmosets as Laboratory Animals .4. Microbiology of Laboratory Kept Marmosets. *Lab Anim Care* **17**, 48–70 (1967).
53. W. M. Koh, *et al.*, The Epidemiology of Hand, Foot and Mouth Disease in Asia: A Systematic Review and Analysis. *Pediatr Infect Dis J* **35**, e285-300 (2016).
54. K. Adachi, A. Muraishi, Y. Seki, K. Yamaki, M. Yoshizuka, Coxsackievirus B3 genomes detected by polymerase chain reaction: evidence of latent persistency in the

- myocardium in experimental murine myocarditis. *Histol Histopathol* **11**, 587–596 (1996).
55. Q. Sun, *et al.*, The specific and essential role of MAVS in antiviral innate immune responses. *Immunity* **24**, 633–642 (2006).
 56. J. P. Wang, *et al.*, MDA5 and MAVS mediate type I interferon responses to coxsackie B virus. *J Virol* **84**, 254–260 (2010).
 57. T. Li, *et al.*, ADAR1 Stimulation by IFN- α Downregulates the Expression of MAVS via RNA Editing to Regulate the Anti-HBV Response. *Mol Ther* **29**, 1335–1348 (2021).
 58. P. P. Liu, J. W. Mason, Advances in the understanding of myocarditis. *Circulation* **104**, 1076–1082 (2001).
 59. L. H. Chow, K. W. Beisel, B. M. McManus, Enteroviral infection of mice with severe combined immunodeficiency. Evidence for direct viral pathogenesis of myocardial injury. *Lab Invest* **66**, 24–31 (1992).
 60. G. Hufnagel, N. Chapman, S. Tracy, A non-cardiovirulent strain of coxsackievirus B3 causes myocarditis in mice with severe combined immunodeficiency syndrome. *European Heart Journal* **16**, 18–19 (1995).
 61. M. A. Opavsky, *et al.*, Susceptibility to myocarditis is dependent on the response of alphabeta T lymphocytes to coxsackieviral infection. *Circ Res* **85**, 551–558 (1999).
 62. S. A. Huber, A. M. Feldman, D. Sartini, Coxsackievirus B3 induces T regulatory cells, which inhibit cardiomyopathy in tumor necrosis factor-alpha transgenic mice. *Circ Res* **99**, 1109–1116 (2006).
 63. K. Li, *et al.*, Differential macrophage polarization in male and female BALB/c mice infected with coxsackievirus B3 defines susceptibility to viral myocarditis. *Circ Res* **105**, 353–364 (2009).
 64. Y. Shi, *et al.*, Regulatory T cells protect mice against coxsackievirus-induced myocarditis through the transforming growth factor beta-coxsackie-adenovirus receptor pathway. *Circulation* **121**, 2624–2634 (2010).
 65. J. R. Lane, D. A. Neumann, A. Lafond-Walker, A. Herskowitz, N. R. Rose, Role of IL-1 and tumor necrosis factor in coxsackie virus-induced autoimmune myocarditis. *J Immunol* **151**, 1682–1690 (1993).
 66. M. Paulsen, *et al.*, MAVS Deficiency Is Associated With a Reduced T Cell Response Upon Secondary RSV Infection in Mice. *Front Immunol* **11**, 572747 (2020).
 67. K.-M. R. Prasad, Y. Xu, Z. Yang, S. T. Acton, B. A. French, Robust cardiomyocyte-specific gene expression following systemic injection of AAV: in vivo gene delivery follows a Poisson distribution. *Gene Ther* **18**, 43–52 (2011).

68. S. E. Jacobs, D. M. Lamson, K. St George, T. J. Walsh, Human rhinoviruses. *Clin Microbiol Rev* **26**, 135–162 (2013).
69. J. M. Greve, *et al.*, The major human rhinovirus receptor is ICAM-1. *Cell* **56**, 839–847 (1989).
70. D. E. Staunton, *et al.*, A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell* **56**, 849–853 (1989).
71. J. E. Tomassini, *et al.*, cDNA cloning reveals that the major group rhinovirus receptor on HeLa cells is intercellular adhesion molecule 1. *Proc Natl Acad Sci U S A* **86**, 4907–4911 (1989).
72. F. Hofer, *et al.*, Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc Natl Acad Sci U S A* **91**, 1839–1842 (1994).
73. J. M. Casasnovas, T. A. Springer, Kinetics and thermodynamics of virus binding to receptor. Studies with rhinovirus, intercellular adhesion molecule-1 (ICAM-1), and surface plasmon resonance. *J Biol Chem* **270**, 13216–13224 (1995).
74. R. Moser, *et al.*, Neutralization of a common cold virus by concatemers of the third ligand binding module of the VLDL-receptor strongly depends on the number of modules. *Virology* **338**, 259–269 (2005).
75. W. C. Au, P. A. Moore, W. Lowther, Y. T. Juang, P. M. Pitha, Identification of a member of the interferon regulatory factor family that binds to the interferon-stimulated response element and activates expression of interferon-induced genes. *Proc Natl Acad Sci U S A* **92**, 11657–11661 (1995).
76. R. Lin, Y. Mamane, J. Hiscott, Structural and functional analysis of interferon regulatory factor 3: localization of the transactivation and autoinhibitory domains. *Mol Cell Biol* **19**, 2465–2474 (1999).
77. S. Liu, *et al.*, Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, aaa2630 (2015).
78. R. B. Seth, L. Sun, C.-K. Ea, Z. J. Chen, Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).
79. A. Antonczyk, *et al.*, Direct Inhibition of IRF-Dependent Transcriptional Regulatory Mechanisms Associated With Disease. *Front Immunol* **10**, 1176 (2019).
80. M. Yoneyama, *et al.*, Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J* **17**, 1087–1095 (1998).

81. Q. Feng, *et al.*, Enterovirus 2Apro Targets MDA5 and MAVS in Infected Cells. *J Virol* **88**, 3369–3378 (2014).
82. A. Mukherjee, *et al.*, The coxsackievirus B 3C protease cleaves MAVS and TRIF to attenuate host type I interferon and apoptotic signaling. *PLoS Pathog* **7**, e1001311 (2011).
83. B. Wang, *et al.*, Enterovirus 71 protease 2Apro targets MAVS to inhibit anti-viral type I interferon responses. *PLoS Pathog* **9**, e1003231 (2013).
84. C. M. Smolko, K. A. Janes, An ultrasensitive fiveplex activity assay for cellular kinases. *Sci Rep* **9**, 19409 (2019).