PNAS
www.pnas.org

Main Manuscript for

Computational Modeling of Protracted HCMV Replication using Genome Substrates and Protein Temporal Profiles

Christopher E. Monti[a,b,d,1], Rebekah L. Mokry[a,1,3], Megan L. Schumacher[a], Ranjan K. Dash[b,c,d,2], and Scott S. Terhune[a,b,d,2]

[a]Department of Microbiology and Immunology, Medical College of Wisconsin, Milwaukee, WI-53226
[b]Department of Biomedical Engineering, Medical College of Wisconsin, Milwaukee, WI-53226
[c]Department of Physiology, Medical College of Wisconsin, Milwaukee, WI-53226
[d]Center of Systems and Molecular Medicine, Medical College of Wisconsin, Milwaukee, WI-53226

Correspondence Authors
Scott S. Terhune
8701 W Watertown Plank Rd, Milwaukee, WI-53226
(414) 955-2511
Email: sterhune@mcw.edu

Ranjan K. Dash
8701 W Watertown Plank Rd, Milwaukee, WI-53226
(414) 955-4497
Email: rdash@mcw.edu

Author Contributions
[1]C.E.M. and R.L.M. contributed equally to this work as first authors.
[2]S.S.T. and R.K.D. contributed equally to this work as senior authors.

[3]Present address: Department of Immunobiology, University of Arizona, Tucson, AZ-85719

Author Contribution Statement
R.L.M. and S.S.T. designed *in vitro* experiments; R.L.M., M.L.S., and C.E.M. performed *in vitro* experiments; R.L.M. and S.S.T. contributed to new reagent preparation and analysis; C.E.M. and R.K.D. developed and tested *in silico* computational models and performed simulations; C.E.M., R.L.M., M.L.S., R.K.D., and S.S.T. analyzed the data; and C.E.M., R.L.M., M.L.S., R.K.D., S.S.T. wrote and edited the manuscript and approved the final submission.

Competing Interest Statement: The authors declare no competing interest.

Classification: Microbiology; Virology; Systems Biology; Computational Biology; Modeling Biological Systems.

Keywords: Human Cytomegalovirus; Computational Modeling; Biological Networks; Viral Replication; Viral Egress.

This PDF file includes:

Main Text

Figures 1 to 4

ABSTRACT

Human cytomegalovirus (HCMV) is a major cause of illness in immunocompromised individuals. The HCMV lytic cycle contributes to the clinical manifestations of infection. The lytic cycle occurs over approximately 96 h in diverse cell types and consists of viral DNA (vDNA) genome replication and temporally distinct expression of hundreds of viral proteins. Given its complexity, understanding this elaborate system can be facilitated by the introduction of mechanistic computational modeling of temporal relationships. Therefore, we developed a multiplicity of infection (MOI)-dependent mechanistic computational model that simulates vDNA kinetics and late lytic replication based on in-house experimental data. The predictive capabilities were established by comparison to *post hoc* experimental data. Computational analysis of combinatorial regulatory mechanisms suggests increasing rates of protein degradation in association with increasing vDNA levels. The model framework also allows expansion to account for additional mechanisms regulating the processes. Simulating vDNA kinetics and the late lytic cycle for a wide range of MOIs yielded several unique observations. These include the presence of saturation behavior at high MOIs, inefficient replication at low MOIs, and a precise range of MOIs in which virus is maximized within a cell type, being 0.382 to 0.688 IU per fibroblast. The predicted saturation kinetics at high MOIs are likely related to the physical limitations of cellular machinery, while inefficient replication at low MOIs may indicate a minimum input material required to facilitate infection. In summary, we have developed and demonstrated the utility of a data-driven and expandable computational model simulating lytic HCMV infection.

SIGNIFICANCE

The complex HCMV lytic replication cycle is associated with the clinical manifestations of HCMV infection. This work uses a novel computational modeling approach based on experimental data to study viral DNA replication, late HCMV protein expression, and infectious virus production. The results demonstrate dynamic relationships and predict a range of MOIs where HCMV replication is most favorable. Introduction of mechanistic modeling reveals new parameters and measurable events required to fully understand the complex interplay between viral and host processes.

78 Ultimately, this quantitative understanding of relationships *in vitro* will lead to quicker development
79 of new monitoring and prophylaxis strategies against HCMV.

INTRODUCTION

Human cytomegalovirus (HCMV) is a betaherpesvirus with an estimated global seropositivity rate of approximately 83% [1]. HCMV is the leading cause of congenital birth defects [2, 3] and a major cause of morbidity and mortality in immunocompromised hosts, especially in hematopoietic stem cell or solid organ transplant patients [4, 5]. Primary infection with HCMV in immunocompetent patients results in a variety of manifestations ranging from asymptomatic infection to a mononucleosis-like syndrome [4]. CMV disease in immunosuppressed transplant patients is defined as CMV infection accompanied by clinical signs and symptoms, and can be broadly categorized into either end-organ CMV disease or CMV syndrome [6]. HCMV is known to have two distinct life cycles, lytic and latent. It is the lytic replication cycle that is associated with the clinical manifestations of CMV syndrome and disease [7].

The HCMV lytic replication cycle has an approximately 96 h duration *in vitro* in fibroblasts that culminates in infectious virions and destruction of the infected cell [7]. Variability exists in length of replication between cell types and is influenced by factors such as mechanism of particle entry and timing of genome delivery to the host nucleus [8-11]. The lytic cycle is marked by viral DNA (vDNA) genome replication, the temporally variant expression of both viral RNAs [12] and viral proteins [13], and the production of new virus. There are over 700 translated open reading frames that have been identified as potential proteins contributing to the HCMV lytic replication cycle [14]. Given this vast number of proteins and the even larger number of permutations of potential protein interactions, the HCMV lytic replication cycle is an extremely complex process.

Computational modeling of biological systems has been utilized extensively in many disciplines. For example, pharmacologists have employed computational models to describe the pharmacokinetic and pharmacodynamic properties of drugs for over 30 years [15-18]. Recently, computational modeling has been applied to study other biological topics such as cell cycle [19-23], viral infections such as hepatitis C [24-26], and early events in HCMV infection [27] as well as HCMV replication in patient samples [28] and drug treatment [29]. Existing models of *in vitro* HCMV lytic replication only focus on the early time points after infection [27] or the ability of

HCMV to alternate between lytic and latent replication cycles [30]. Currently, there are no models describing the events leading to production of infectious virions.

In this work, we have developed an empirical model of intracellular viral genome (i.e., vDNA) replication and then utilized the output from this model as input to drive a mechanistic computational model of the late viral protein temporal class expression and viral egress. Each of these models was developed based on experimental data obtained at several multiplicities of infection (MOIs) in fibroblasts and then compared to experimental data obtained *post hoc* for model validation. Using *in vitro* and *in silco* experiments, we have elucidated a range of MOIs where both vDNA and cell-free virus production are maximized in infected fibroblasts. Our studies predict a minimum MOI in fibroblasts, below which both replicated vDNA and cell-free virus are less than the initial input. Our studies also demonstrate saturation kinetics where the maximal capacity of cells has been reached. The resulting computational model provides a mechanistic framework on which to build out the many complex relationships, both intra- and intercellular, occurring during HCMV infection and to test complex hypotheses relating multivariate interactions *in silico*.

RESULTS

Inherent limits of efficient virus genome replication kinetics framed by upper and lower thresholds.

HCMV replication occurs via a coordinated and temporal series of events all requiring vDNA. The degree of coordination between replication components has yet to be fully defined due to the exceptional complexity of the viral life cycle. To generate a predictive computational model focusing on late viral events, it was necessary to first generate an empirical model of vDNA synthesis using in-house HCMV vDNA experimental data sets to subsequently utilize as a driving input for a model of the late lytic replication cycle. vDNA synthesis begins as early as 24 hours post infection (hpi) *in vitro* which is influenced by cell type and mechanisms of virion entry [31]. To develop this empirical model of vDNA kinetics, we formulated a simple schematic of different vDNA species that exist during infection (Fig. 1A). This schematic includes input viral genomes

136 termed $vDNA_{in}$, which associates with the target cells. We postulate the existence of genome loss
137 or degradation (characterized by the rate constant k_d) due to both failure of some copies to reach
138 the nucleus and consumption of the vDNA by semi-conservative replication to generate newly
139 synthesized vDNA genome copies. Once vDNA synthesis is initiated, the concentration of
140 replicated vDNA, $vDNA_{rep}$, will begin to increase over time irrespective of genome replication
141 mechanisms (e.g., semi-conservative, rolling circle, homologous recombination, etc.). The sum of
142 $vDNA_{in}$ and $vDNA_{rep}$ is the total cell-associated vDNA ($vDNA_{tot}$), which we can experimentally
143 measure. For the purpose of this empirical model, we set free $vDNA_{tot}$ to be in excess compared
144 to genomes packaged into particles destined to leave the infected cell [32-37]. To develop the
145 empirical model, we experimentally quantified HCMV $vDNA_{tot}$ using different input MOIs (IU/cell)
146 over 96 h (Fig. 1B). We infected confluent MRC-5 fibroblasts using recombinant HCMV from
147 strain TB40/E expressing late protein pp28 in-frame with the fluorescent protein mCherry and IE2
148 in-frame with a cleavable eGFP (IE2-2A-eGFP UL99-mCh). We used absolute genome standards
149 for both HCMV and host cells allowing for comparison between conditions (Fig. 1B). We
150 determined MOIs of 0.1, 0.5, and 5 IU/cell resulted in average viral genomes/cell at 2 hpi
151 ($vDNA_{in,0}$) of 3 ± 1, 13 ± 3, and 131 ± 70, respectively, in this experimental system (Fig. 1C).
152 These data collected from varying times and inputs resulted in the empirical model of $vDNA_{tot}$
153 dynamics shown in Eq. 1-3 (see Methods).

154 The empirical model was fit to each MOI-dependent data independently and nonlinear
155 regression to Eq. 1-3 was performed on the individual estimates to generate $vDNA_{in,0}$-dependent
156 parameters $vDNA_{rep,max}$, t_{50}, and n (Fig. S1). We generated a conversion between MOI and input
157 viral genomes (Fig. 1C), noting that the MOI measurement depends on the method used for
158 titering viral concentrations and number of cells, while quantifying genomes is a universal
159 standard [38]. This conversion ensures unit consistency between output (genomes/cell) and input
160 parameters. Since correspondence between the model and data (Fig. 1D) was strong, we
161 simulated $vDNA_{tot}$ dynamics for varying input $vDNA_{in,0}$ (Fig. 1E). To test the predictive nature of
162 our empirical model of $vDNA_{tot}$, we repeated the experiment using two additional MOIs of 0.01
163 and 0.23 IU/cell, which corresponded to $vDNA_{in,0}$ of 0.2 ± 0.1 and 6 ± 2 genomes/cell,

164 respectively. Model predictions showed good correspondence with the experimental data (Fig. 1E
165 and 1F), corroborating the model. We also generated 3D plots of $vDNA_{tot}$ and $vDNA_{tot}/vDNA_{in,0}$
166 vs. $vDNA_{in,0}$ and time showing the dynamic relationships occurring during infection (Fig. 1G).
167 Separating $vDNA_{tot}$ into its constituents, the change in the concentration for $vDNA_{in}$ was
168 determined by MOI and the decay rate constant k_d, while the dynamics of $vDNA_{rep}$ were
169 determined by the empirical model based on our experimental data (Fig. 1H). Our simulation
170 predicts a maximum increase in $vDNA_{tot}$ of 2.6 logs occurring when $vDNA_{in,0}$ is in the range of 3 to
171 32 genomes/cell, representing MOIs of 0.1 to 1.2 IU/cell in primary fibroblasts using our infection
172 conditions (Fig. 1G). Below this range, the model predicts limited vDNA synthesis; while above
173 this range, we observe saturation kinetics. We hypothesize that this range is the result of intrinsic
174 features of the host cells, namely the influence of entry and intrinsic antiviral responses at lower
175 inputs [31, 39, 40] and potentially related to maximum metabolic or structural capacity of these
176 cells to support replication at higher inputs [41-43].
177
178 **Formulation of a MOI-dependent mechanistic computational model of HCMV late lytic**
179 **replication cycle and associated in-house experimental data.**
180 HCMV virion production involves both a nuclear phase and a cytoplasmic phase. Viral
181 proteins participating in nuclear and cytoplasmic virion production were predominantly
182 categorized as “temporal profile 5” (Tp5) class proteins by Weekes *et al*. [13]. Tp5 class proteins
183 production was determined to depend on $vDNA_{tot}$ [13]. Using our empirical $vDNA_{tot}$ model as a
184 driving input allowing for the estimation of $vDNA_{tot}$ at any time and $vDNA_{in,0}$, we formulated a
185 conceptional relationship (solid lines) and putative regulation mechanisms (dotted lines) for the
186 nuclear and cytoplasmic phases using the nuclear and cytoplasmic proteins $Tp5_1$ and $Tp5_2$,
187 respectively (Fig. 2A).
188 We hypothesize that the level of $Tp5_1$ is dependent on $vDNA_{tot}$ and is influenced by rates
189 of synthesis ($k_{s,1}$) and degradation ($k_{d,1}$) (Fig. 2A). For production of infectious virus, we postulate
190 that $vDNA_{tot}$ must first associate with $Tp5_1$ class proteins ($k_{s,C}$) eventually resulting in a capsid
191 containing a single genome. To obtain an experimental baseline for our hypotheses, we

measured a representative HCMV nuclear $Tp5_1$ class protein, pUL44 (52 kDa), using immunoblot analysis with a standard curve (Fig. 2B), completed in parallel with $vDNA_{tot}$ measurements (Fig. 1B). The standard curve consists of dilutions of whole cell lysates following infection at 96 hpi and used in the quantification process for comparison between blots. These standards were used for each antibody and immunoblot allowing us to compare relative signal intensities between experiments (Fig. 2C). Our analysis resulted in the relative expression level of total $Tp5_1$ over time. Additionally, at each time point, our analysis resulted in $Tp5_1$ expression levels relative to each $vDNA_{in,0}$ (Fig. 2D). We included the quantification by mass spectrometry for pUL44 levels by Weekes *et al.* [13] showing comparable expression kinetics at the highest MOI of 131 genomes/cell (Fig. 2D).

The second phase of the lytic replication cycle involves the egress of vDNA-containing capsids into a cytoplasmic assembly compartment and associating with a second set of Tp5 proteins, $Tp5_2$ (Fig. 2A). Similar to $Tp5_1$, we hypothesize that $Tp5_2$ is dependent on $vDNA_{tot}$ and is influenced by rates of synthesis ($k_{s,2}$) and degradation ($k_{d,2}$). For production of infectious virus, capsids associate with $Tp5_2$ class proteins ($k_{s,P}$) eventually resulting in intracellular particles. We measured the total cellular levels of a representative HCMV cytoplasmic $Tp5_2$ protein, pp28 (28 kDa pp28; 56 kDa pp28-mCherry) using immunoblot analysis and an antibody against pp28 (Fig. 2E). Using a standard curve (Fig. 2F), we quantified the relative expression of $Tp5_2$ over time and in proportion to $vDNA_{in,0}$ (Fig. 2G), and included the relative levels measured by Weekes *et al.* [13], again showing nearly identical expression kinetics at the highest MOIs.

Productive viral replication results in the release of infectious, cell-free virus (Fig. 2A). To experimentally quantify this phenomenon, we measured viral titers starting at 24 hpi in culture media from HCMV infections at the average $vDNA_{in,0}$ of 3 (MOI 0.1), 14 (MOI 0.5), and 131 (MOI 5) genomes/cell (Fig. 2H). Titers of the time course media were determined by quantifying the resulting HCMV IE1-positive cells in a new culture and defining infectious units per ml (IU/ml). We observed titers at 24 hpi in proportion to inputs with similar fold increases for $vDNA_{in,0}$ of 14 and 3 genomes/cell by 96 hpi (Fig. 2H). In contrast, $vDNA_{in,0}$ of 131 genomes/cell exhibited saturation kinetics with titers at 96 hpi comparable to $vDNA_{in,0}$ of 14 genomes/cell which supports our

previous observation that the culture has a maximal capacity. We observed relatively high levels of infectious virus present at 24 hpi (Fig. 2H), and we speculate that this is residual inoculum as it exhibited poor infectivity compared to the input stock and 72-96 hpi cell-free virus (Fig. 2I). Based on this information, we elected to use the resulting titers from time points between 48 and 96 hpi for subsequent modeling studies.

Identification of data-driven mechanistic computational model of HCMV late lytic replication cycle using an ODE framework and predictive outcomes of the model.

Because of the complexity and inclusion of multiple biological processes within a simple computational model of late lytic replication cycle and the observation that several data points plateau late in infection suggesting additional regulations, we tested six competing models (see SI Appendix) which differed at the level of $Tp5_1$ and $Tp5_2$ regulation (Fig. 2A) to define the minimal model (Fig. 3A) that best describes the data. Using mass balance, we derived coupled, nonlinear ordinary differential equations (ODEs) describing changes in $Tp5_1$ and $Tp5_2$ classes of proteins and its complexes and virus production using the basic frameworks in Figs. S2-S7. We estimated unknown parameters of each model by fitting model solutions of the relevant state variables to total $Tp5_1$ and $Tp5_2$ data (Figs. 2D and 2G) and extracellular virus data (Fig. 2H) using a pseudo-Monte Carlo minimization parameter estimation protocol. Estimated parameter values for each of our models can be found in Figs S2C-S7C. Pseudo-Monte Carlo fits of each model to the data from Fig. 2 are shown in Figs. S8 and S9. For further analysis, we opted to use the average parameter sets (Fig. S9, dashed curves) since they were based on the average of 200 pseudo-Monte Carlo fittings, reducing the risk of presenting an erroneous parameter set due to random chance of minimization of the sum of squares of errors (SSE) objective function.

After parameter estimation for each model, we then began the process of model selection (see SI Methods) to identify the optimal model. We performed two statistical tests, the Akaike Information Criteria (AIC) and F-test [44] (Figs. S10 and S11), to determine which of the models most accurately describes the experimental Tp5 protein level (Fig. 2D and 2G) and virus production over time (Fig. 2H) data for all $vDNA_{in,0}$. We then analyzed the contribution of each

248 regulatory component (Fig. S12) and the variability of each estimated parameter value, SSE, and
249 model fit (Figs. S2-S7) to help confirm the predictions made by the statistical tests and increase
250 our confidence in the model selection.
251 Results of the AIC comparison between different models clearly showed that model 3 has
252 the highest likelihood of correctness, followed closely by model 1 (Fig. S10). Model 1 was less
253 than 90% likely only when compared to models 2 and 3 (Fig. S10). The top row of Fig. S11,
254 showing the comparison of the least complex model 1 tested against all other, more complex
255 models using the F-test, indicates that only model 3 yields an F-statistic and corresponding p-
256 value lower than our Bonferroni-corrected significance level of 6×10^{-3}, which was determined *a*
257 *priori*. Furthermore, when the one-regulation models (2, 3, and 4) were compared against the two
258 regulation models (5 and 6), only the comparison between models 4 and 5 led to a significant p-
259 value. Since models 4 and 5 are not distinguishable, these results indicate that a one-regulation
260 model is likely optimal. In summary, the results of both the AIC and F-test dictate that model 3 is
261 likely the optimal model to describe the data (Figs. S2C-S7C and S10-S11).
262 When analyzing at the contribution of each regulatory parameter (Fig. S12), we can see
263 that the regulation present in model 3 shows a smooth curve spanning the entire range of the
264 function (e.g., 0-1) over its domain for all $vDNA_{in,0}$ (Fig. S12B). Conversely, the regulation in
265 model 2, for example, shows little change in the regulatory function and its effect is only
266 appreciable at high $vDNA_{in,0}$ and late in the infection (Fig. S12A). This trend is mimicked by model
267 4 (Fig. S12C) and by the feedback inhibition terms ($R_{1,6}$) in model 6 (Fig. S12E). Interestingly, if
268 $R_{1,6}$ is reduced to unity for both $Tp5_1$ and $Tp5_2$ $R_{1,6}$ terms, the ODEs for model 6 reduce to model
269 3. Furthermore, the $R_{2,6}$ terms in model 6 show a range similar to that exhibited by model 3 (Fig.
270 S12C and S12E). Finally, model 5 shows inconsistency in its regulatory terms: $R_{1,5}$ for Tp_2, but
271 not $Tp5_1$, shows acceptable range, while $R_{2,5}$ shows acceptable range for $Tp5_1$, but not $Tp5_2$. This
272 inconsistency is likely due to numerical compensation, where an increase in one estimated
273 parameter can be compensated by a decrease in a conjugate estimated parameter leading to
274 many equivalent solutions. This suggests that there is no unique solution for the parameters
275 associated with this model, decreasing our confidence in this model's true correctness. In support

of this argument, Figs. S2C-7C show a large variability in the majority of parameters for models 2, 4, 5, and 6. This is also present in Figs. S8 and S9. In Fig. S8, a large spread of the 200 pseudo-Monte Carlo iterations indicates a large variability in the estimated parameter values, and in Fig. S9, a large difference between the dashed and solid curve indicates a large difference between the average parameter set and the parameter set yielding the lowest SSE. This same result can also be seen in the variability between SSE obtained with the average parameter set and the lowest obtained SSE and AIC (Figs. S2C-S7C). Interestingly, model 3 not only has a small variability in obtained parameter values (Fig. S4C), but its average and minimum values are so close that they are indistinguishable in Fig. S9. This fact is supported by the SSE and AIC values shown in Fig. S4C.

In summary, to determine the most optimal model, we implemented a careful statistical analysis to determine the optimal number of regulatory terms as well as integrative examination of the AIC, contribution of regulatory terms, parameter/model fit variability, and SSE variability. Results of this analysis led to the conclusion that, in fact, model 3 yields the most optimal and parsimonious description of the data.

In the best-fitting model (Fig. 3A), $vDNA_{tot}$ drives $Tp5_1$ protein production ($k_{s,1}$, $K_{m,1}$), which combines with $vDNA_{tot}$ to generate capsids ($k_{s,C}$) as formulated in Eqs. 5 and 7 (see Methods). Regulation of $Tp5_1$ production occurs through acceleration of $Tp5_1$ degradation ($K_{m,3}$). In this model, we hypothesize that $vDNA_{tot}$ is in excess due to the large size of the nuclear replication center and production of concatemeric genomes with egress requiring single genome-containing capsids [45]. Additionally, it is known that procapsids are rapidly converted to vDNA-containing capsids suggesting that vDNA must be in excess for the process to be kinetically favorable [46]. We hypothesize that $Tp5_1$ consumption is through normal cellular degradation pathways ($k_{d,1}$) and capsid assembly ($k_{s,C}$). Although pUL44 is not a capsid protein, we used measurements of pUL44 levels as a representative nuclear protein with Tp5 kinetics [13, 47, 48]. This $Tp5_1$ placeholder will be expanded to account for additional proteins in future studies. In our model, $Tp5_2$ production ($k_{s,2}$, $K_{m,2}$) represents late cytoplasmic proteins that associate with the capsid after nuclear egress. We hypothesize that these proteins are consumed through

degradation pathways ($k_{d,2}$) and particle assembly ($k_{s,P}$) as accounted for in Eqs. 6 and 8 (see Methods). Regulation of $Tp5_2$ production occurs through acceleration of $Tp5_2$ degradation ($K_{m,4}$). Intracellular viral particles leave the cell (k_{ex}) and their concentrations are diluted in culture media as accounted for in Eq. 9 (see Methods) (Fig. 3A). We normalized experimental data to the maximum of the dataset with $vDNA_{in,0}$ = 131 genomes/cell at 96 hpi to maintain magnitude consistency in the SSE objective functions used in parameter estimation.

Fit of the model 3 system of ODEs (Eqs. 5-9) to experimental data is shown in Figs. 3B, 3D, and 3F. Overall model 3 fit to experimental data was acceptable giving an AIC of -230.5 (Fig. S4C). Simulations of $Tp5_1$ and $Tp5_2$ expression as well as virus production are shown in 2-dimensional and 3-dimensional plots over time and many $vDNA_{in,0}$ in Figs. 3C, 3E, and 3G. Simulating changes in Tp5 levels upon increasing MOI (or $vDNA_{in,0}$) showed expression starting to occur at immediate early times (Fig. 3C and 3E). Starting at 24 hpi, we simulated infectious HCMV production at varying $vDNA_{in,0}$ (Fig. 3G). Predicted virus production kinetics showed saturation at high $vDNA_{in,0}$ and suboptimal replication at low $vDNA_{in,0}$.

To assess the correlation and individual estimability of each parameter for model 3, we generated a correlation matrix (Fig. S13A) and performed a sensitivity analysis (Fig. S13B) [49-51]. The correlation matrix showed that there was a high degree of correlation between the protein degradation rate constants k_d's and other parameters within the same ODEs (Fig. S13A). The sensitivity analysis (Fig. S13B), which shows the relative change in parameter value (p/p_0) versus the relative change in the error function (SSE/SSE_0), showed expected parabolic behavior for all parameters but k_{ex}. Regulatory parameters $K_{m,3}$ and $K_{m,4}$ do show parabolic behavior between $0.5p_0$ and $1.5p_0$, albeit on a much smaller scale than $K_{m,1}$ and $K_{m,2}$. To break parameter correlation and minimize the number of estimated parameters, we opted to fix k_{d1}, k_{d2}, and k_{ex} in all models for parameter estimation and further model analyses. Simulations of individual model state variables in absolute units for model 3 can be found in Fig. S14.

Model of HCMV late lytic replication cycle predicts conditions for maximal efficiency for HCMV replication.

Viral titers are experimentally measured in absolute quantities such as IU/ml. In order to revert back to these units, the simulated data were vertically scaled by a factor of 10^7 IU/ml as shown in Fig. 4A comparing $vDNA_{in,0}$ (MOI) and total virus production over time. This multiplicative factor was chosen since it was the order of magnitude of the normalizing factor (i.e., maximum of the virus titer data set) used to model virus data in Fig. 3F. We have extended the range of each variable and, using numerical estimates, we defined the approximate range containing the maximum fold change for $vDNA_{tot}$ occurs upon infection $vDNA_{in,0}$ of 9 and 13 genomes/cell (0.688 to 0.994 IU/cell) (Fig. 4B Left). This range is just above that of the extracellular virus which occurs upon infection with 5 to 9 genomes/cell (0.382 to 0.688 IU/cell) in primary fibroblasts (Fig. 4B Right).

Our studies have resulted in a simulation of dynamic relationships occurring during HCMV infection for the purpose of predicting how changing one or more variables will impact others in the complex process of HCMV replication. We plotted the relationship between $vDNA_{tot}$ and relative $Tp5_1$ and $Tp5_2$ levels over time (Figs. 4C and 4D, respectively) and colorized with fold change in vDNA between 2 hpi and 96 hpi. Fig. 2C and 2F suggest an approximate linear range of relative protein level between 0.25 and 1 relative units [52]. Simulations in Figs. 4C and 4D suggest that relative $Tp5_1$ and $Tp5_2$ quantities only begin to reach levels within this linear range at 96 hpi when the fold change in vDNA between 2 hpi and 96 hpi is maximized (Figs. 4C and 4D, red area). Furthermore, relative levels of virus production within this same range of $vDNA_{in,0}$ are within approximately 1×10^{-4} to 0.83 which is the largest acceleration in virus production (Fig. 4E). Hence, we hypothesize that a pattern of a $Tp5_1$ (e.g., pUL44) or $Tp5_2$ protein (e.g., pp28) kinetic expression similar to that in the red region in Figs. 4C or 4D would indicate a maximally efficient virus production. It is important to note in Figs. 4C and 4D that at very high $vDNA_{in,0}$, there is an increase in $Tp5_1$ and $Tp5_2$ proteins prior to 24 hpi and is hypothesized to be a result of dysregulated viral gene expression kinetics. In support of this hypothesis, simulations Fig. 4E also support inefficient virus production at high $vDNA_{in,0}$.

Recombinant HCMV strains containing tagged viral proteins are routinely used to define expression and function of viral proteins during replication. To obtain higher temporal resolution

data during the 4-day replication cycle and evaluate the accuracy of simulated data built from limited time points, we infected fibroblasts with recombinant HCMV TB40/E IE2-2A-eGFP UL99-mCh at multiple $vDNA_{in,0}$, as described previously. Using this system, we captured fluorescence data for pp28-mCherry ($Tp5_2$) along with free eGFP every 2 h for 97 h using a live cell imaging platform (Fig. 4F and movies S1-S3). We determined the average relative mCherry signal intensity to maximal signal occurring at infection of $vDNA_{in,0}$ = 131 genomes/cell at 97 hpi (Fig. 4H). Fig. 4G shows single cell analysis of the $vDNA_{in,0}$ = 3 genomes/cell fluorescence data, highlighting the cell-to-cell variability that is frequently lost when looking at data from infected cells pooled for traditional protein analysis (e.g., immunoblot). Fig. 4I shows predicted $Tp5_2$ expression from model 3. In a *post hoc* comparison of Figs. 4H (in vitro data) and 4I (in silico model), we can see good correspondence of the model and the data indicating that our simulation of pp28 closely matches data obtained from an alternative experimental measurement and at a higher temporal resolution that is not possible using standard methods of protein quantification.

DISCUSSION

HCMV lytic replication is an immensely complex process that occurs over a relatively protracted time period and involves hundreds of viral and host proteins in an elaborate interplay that eventually results in newly produced virions. As such, it is infeasible to attempt to quantitatively understand the process without the use of computational aids. Computational modeling is a well-established tool while its application to the HCMV lytic replication cycle remains relatively novel. In our studies, we developed two models based on experimental data. The first was an empirical model of vDNA replication (Fig. 1). In this model, we hypothesized an initial decay of $vDNA_{in}$, which was then followed by an increase in $vDNA_{rep}$. The output from this vDNA model was then used to drive a model of the late lytic replication cycle, which began with late viral protein expression and culminated in predictions of capsid, intracellular viral particle, and extracellular virus production kinetics (Fig. 3). The model of viral protein expression predicted an increase in protein degradation at late times that followed the trend of increasing vDNA as infection progressed. This interaction could be related to an increase in proteasome activity that

promotes protein degradation and demonstrated to occur for HCMV [53-55], lending further experimental support to this proposed mechanism. In addition, herpesvirus capsids do undergo protease-dependent maturation [56], and numerous DNA and RNA viruses undergo late-stage maturation events involving protein cleavage, most notable is HIV [57]. The HIV-1 protease inhibitor, Nelfinavir disrupts secondary HSV-1 envelopment [57], and its anti-herpesvirus efficacy is under clinical investigations. Our simulations have uncovered a possible role for increased protease activity in HCMV maturation.

We observed several significant predictions from this data-driven computational model: (1) saturation kinetics at high MOIs, (2) inefficient replication at low MOIs, and (3) a range of MOIs where virus replication is maximized in primary human fibroblasts. Results shown in Fig. 3 and 4 demonstrate an ideal MOI where vDNA and virus production are maximized. MOIs below this maximal range yield suboptimal replication efficiency and are predicted by our model to lead to abortive infections. Furthermore, MOIs above this range represent diminishing returns. This is an example of a biological Goldilocks phenomenon [58], where both too little and too much virus applied to a system leads to suboptimal replication. Another common analysis performed on models involves determination of a rate-limiting step. In the setup of our model, the forward synthesis rates ultimately leading to infectious virus production are $k_{s,1}$, $k_{s,2}$, $k_{s,C}$, $k_{s,P}$, and k_{ex}, and their values are shown in the SI Appendix. While it is likely impossible to define an overall rate-limiting step for all HCMV replication from this model, we can potentially define rate-limiting steps in several subprocesses included in our model. First, we found that the slowest rate of protein production was $k_{s,1}$ (6.24 x 10^2 genomes/cell/h), indicating that $Tp5_1$ proteins are produced slower than $Tp5_2$ proteins (7.6 x 10^2). Second, we found that $k_{s,C}$ (1.116 x 10^{-6} cell/genomes/h) was smaller than $k_{s,P}$ (3 .090 x 10^{-5} cell/genomes/h) indicating that formation of intranuclear, loaded capsids was the rate-limiting step in the process of capsid synthesis and viral egress, which has been previously suggested in studies on nuclear egress [59].

Our data-driven models of viral genome synthesis and late protein expression can be expanded to include other mechanistic components of the replication cycle. For example, the model by Vardi et al. [27] predicts expression kinetics via the major immediate early promoter

(MIEP) and feedforward activation of the IE1 protein based on virion-delivered pp71 (Tp5) [13, 60, 61]. It is reasonable to propose that the MIEP-dependent constants may, in fact, be non-constant and vary with MIEP containing-$vDNA_{in,0}$. Given that at saturating MOIs most cells are multiply infected, the parameter representing basal IE1 expression independent of transactivation and feedforward mechanisms could be of greater influence on the ODE governing IE1 expression. It is likely that as $vDNA_{in,0}$ increases, the concentration of pp71 will also increase leading to early saturation of its abilities to de-silence the MIEP. In fact, it is known that the effectiveness of pp71 wanes as MOI increases to the extent that it is only required for infection at low MOI [62]. The Vardi et al. [27] model predicts that increasing pp71 concentrations could sustain IE1 expression even in the absence of positive feedback. Thus, a reasonable hypothesis resulting from combining models is that despite low production of $Tp5_2$ proteins at low MOIs, supplementation with excess pp71 during subsequent infection could sustain IE1 expression as predicted by Vardi et al. [27] and, by extension, a productive infection despite a potentially suboptimal or even abortive MOI as seen in our simulations. To some degree, this has been demonstrated by the inclusion of an expression vector for pp71 during the process of obtaining infectious virus from transfected genomes [63].

To connect additional processes, substantial amounts of published kinetic data exist that can be used to build additional empirical or ODE-based models by simply aligning genome kinetics to our simulation. For our studies, we specifically used absolute quantification of viral genomes/cell to avoid discrepancies introduced when using an MOI-based approach. As an example, we overlaid protein expression data from Weekes et al. [13] showing near identical kinetics of pUL44 and pp28 at $vDNA_{in,0}$ of 131 genomes/cell. This alignment allows for expansion to multiple expression classes and perhaps specific viral proteins. Alternatively, tracking recombinant viruses expressing fluorescently tagged viral proteins as done here provides a unique opportunity to obtain higher resolution kinetic data. Recently Rand et al. [64] introduced a triple fluorescent HCMV strain with fluorescence in each of three expression classes. As we move forward, we anticipate using this new base model to account for more precise mechanisms governing HCMV replication.

While our studies present a robust and predictive model of late HCMV replication, there are some limitations. Our experimental methods used a single cell type, a single strain of HCMV, a single stock of virus, and growth arrested cells. These steps were necessary to reduce the complexity of the system and potentially control the variability in the data, both of which are necessary to facilitate computational modeling. Imposing these experimental restrictions, however, reduces the generalizability of our model. Future studies will explore kinetics in different cell types and in a steady-state infection as recently done in hepatitis C [24]. The creation of an ODE-based model also required the introduction of simplifying assumptions. First, we assumed that the $vDNA_{tot}$ in the system was a good approximation of the free vDNA available for protein expression and gene regulation. This assumption was required so that we were able to use the $vDNA_{tot}$ value at any time and $vDNA_{in,0}$ as the driving force for the model of the late lytic replication cycle (Fig. 3). This assumption is justified because mature C-type capsids represent a small fraction of the total capsid types in the nucleus with viral terminase activity requiring an excess of vDNA templates for packaging [32-37]. Next, the mechanisms proposed in our model of the late lytic replication cycle represent the lumping of many, potentially unmeasurable, smaller subprocesses. For instance, vDNA goes through an mRNA intermediary to produce viral proteins. We accounted for the mechanistic kinetics of many binding events during protein expression and the potential of protein synthesis machinery (e.g., ribosomes) saturation by introducing Michaelis-Menten kinetics, characterized by K_m parameters. We introduced these kinetics to avoid using delay differential equations (DDEs) [65], which generally are slower to solve and not amenable to parameter estimation where the model equations are solved many times and compared to experimental data via the SSE in order to obtain optimal parameters. Simplifying assumptions were made in an effort to avoid overparameterization, which is a problem with many mathematical models, including ODE-based models [65]. Future research will focus on obtaining viral mRNA kinetic data as well as higher resolution kinetic data for late proteins that may be used to relax the aforementioned assumptions. The advantage of using a computational model such as the one presented in this article is that it can be expanded to explicitly account for the exact pathway once it has been elucidated in full. For now, however, we can model the relationship as

presented and still generate useful predictions. Finally, the Bonferroni-corrected p-value used when determining significance for the F-test may be too stringent for the number of comparisons that were made. Given the nature of our work, however, we were inclined to strictly minimize the SSE.

We used a systematic and modular approach to modeling in an effort to provide an accurate and robust mathematical representation of the complex lytic replication cycle. First, we employed a pseudo-Monte Carlo parameter estimation protocol using the minimal number of estimable parameters and fixing other highly correlated or insensitive parameters to obtain an optimal parameter set that minimized the sum squared error for the model of the late lytic replication cycle (Fig. 3). For model selection we performed three analyses: (1) As described in the results, we used a systematic approach involving the AIC and F-test to justify or reject the inclusion of additional parameters and avoid overparameterization. This led to the conclusion that model 3 was the best-fitting model; (2) We analyzed the contribution of each regulatory component in each model and the variability of estimated parameters and SSE. If estimated parameters led to small contributions of regulatory components or showed high variability, that model was rejected, as described in the results; and (3) Finally, the results of a runs test on model 3 [44] failed to reject the null hypothesis indicating that the curve does not systematically deviate from the data. Admittedly, it is possible that a different combination of mechanisms or a completely different model may, in fact, provide a more accurate description of experimental data and might not have been considered. However, we believe that the combination of these methods should provide a robust and parsimonious model.

MATERIALS AND METHODS

Cells, Viruses, and Biological Reagents

Dual fluorescently tagged TB40/E HCMV expressing IE2-2A-eGFP and UL99-mCherry was generously provided by Dr. Eain Murphy. Viral stocks were propagated as a P1 stock on MRC-5 fibroblasts (ATCC) and concentrated by collecting culture medium and pelleting through a sorbitol cushion (see SI Methods). Viral stock titers were obtained by a limiting dilution assay

(TCID$_{50}$) assay on MRC-5s. For studies involving infected cells, MRC-5 fibroblasts were plated onto 6-well dishes at a density of approximately 300,000-500,000 cells/well and allowed to grow until confluent and growth arrested for at least two days for cell cycle synchronization. Cells were infected at the indicated MOI using an approximation of 1 x 10^6 cells per confluent well. Further information regarding titering as well as protein and nucleic acid assays can be found in SI Methods.

Model Development, Parameterization, Validation, and Statistical Testing

Empiric Model of vDNA Kinetics. For vDNA replication, we developed the following empirical model (Eqs. 1-3) based on the hypothesized schematic in Fig. 1A.

$$vDNA_{tot}(t) = vDNA_{in}(t) + vDNA_{rep}(t) \quad [\mathbf{1}]$$

$$vDNA_{in}(t) = vDNA_{in,0} \cdot e^{-0.1.t}; \; vDNA_{rep}(t) = \frac{vDNA_{rep,max} \cdot t^n}{t_{50}^n + t^n} \quad [\mathbf{2}]$$

where

$$vDNA_{rep,max} = \frac{1.9E4 \cdot vDNA_{in,0}^2}{10.5^2 + vDNA_{in,0}^2}; \; t_{50} = \frac{20 \cdot vDNA_{in,0}}{1.3 + vDNA_{in,0}} + 76.3 \; ; \; n = 1.7e^{-0.001 \cdot vDNA_{in,0}} + 3.1 \quad [\mathbf{3}]$$

vDNA$_{tot}$ is the total vDNA in the system, while vDNA$_{in}$ is the input vDNA contained within the inoculum, and vDNA$_{rep}$ is the newly produced (replicated) vDNA. vDNA$_{in}$ is estimated to have an initial value vDNA$_{in,0}$ linearly proportional to the MOI (Fig. 1C) and to decay over time as a single exponential function with the decay rate constant k_d = 0.1. In the empirical model of vDNA$_{rep}$, vDNA$_{rep,max}$ represents the maximal replication achieved at a specific vDNA$_{in,0}$ or MOI; t_{50} is the horizontal shift component of vDNA$_{rep}$ corresponding to the time required to achieve 50% of maximal replication at a specific vDNA$_{in,0}$ or MOI; and n is the Hill coefficient for replication at a specific vDNA$_{in,0}$ or MOI indicating the degree of effective cooperativity.

The qPCR data in Fig. 1D was used to parameterize the model for each vDNA$_{in,0}$ or MOI independently employing a pseudo-Monte Carlo parameter estimation method described below using several iterations. Parameter vs. vDNA$_{in,0}$ data was gathered and then subjected to nonlinear regression using the MATLAB (MathWorks Inc.) Curve Fitting Tool to generate the parameters for the above equations. These vDNA$_{in,0}$-dependent curves were then input into the parameters for Eq. 2. The pseudo-Monte Carlo parameter estimation protocol for fitting the model

equations involved minimizing the SSE (Eq. 4) using the MATLAB Optimization Toolbox "fmincon" function (see SI Methods). This procedure was run for several iterations, and the parameter set yielding the lowest SSE was selected as the optimal parameter set. For the vDNA model, the SSE was defined as

$$SSE = \sum\left(Data(t) - Model(t)\right)^2. \quad [4]$$

where the data value was the mean of three biological replicates.

Deterministic Model of the Late Lytic Replication Cycle. To develop the mechanistic computational model of late viral replication, we postulated six different models accounting for different regulatory mechanisms as described in Fig. 2A and in the Results section. A general equation for each of the models is described in Eqs. 5-9:

$$\frac{d[Tp5_1]}{dt} = k_{s,1}\frac{[vDNA_{tot}]}{K_m+[vDNA_{tot}]}R_{1,i}(t) - k_{d,1}[Tp5_1]R_{2,i}(t) - k_{s,C}[vDNA_{tot}][Tp5_1] \quad [5]$$

$$\frac{d[Tp5_2]}{dt} = k_{s,2}\frac{[vDNA_{tot}]}{K_m+[vDNA_{tot}]}R_{1,i}(t) - k_{d,2}[Tp5_2]R_{2,i}(t) - k_{s,P}[Capsid][Tp5_2] \quad [6]$$

$$\frac{d[Capsid]}{dt} = k_{s,C}[vDNA_{tot}][Tp5_1] - k_{s,P}[Tp5_2][Capsid] \quad [7]$$

$$\frac{d[Particle]}{dt} = k_{s,P}[Tp5_2][Capsid] - k_{ex}[Particle] \quad [8]$$

$$\frac{d[Virus]}{dt} = k_{ex}\frac{V_{cell}}{V_{media}}[Particle] \quad [9]$$

where [X] represents the concentration of a state variable "X" in the absolute units of vDNA (genomes/cell). $R_{1,i}(t)$ and $R_{2,i}(t)$ represent the putative regulatory components investigated in models 2-6 and are described by:

$$R_{1,i}(t) = \begin{cases} 1 & i = 1,2,3 \\ \frac{K_m}{K_m+[Tp5_{1\text{ or }2}]} & i = 4,5,6 \end{cases} \quad [10]$$

$$R_{2,i}(t) = \begin{cases} 1 & i = 1,4 \\ \frac{K_m}{K_m+[vDNA_{tot}]} & i = 2,5 \\ \frac{[vDNA_{tot}]}{K_m+[vDNA_{tot}]} & i = 3,6 \end{cases} \quad [11]$$

Definitions and values of the parameters in Eqs. 5-9 can be found in the SI Appendix. Nuclear-localized, capsid-forming $Tp5_1$ proteins are produced from $vDNA_{tot}$ (substrate) with a rate $k_{s,1}$ (genomes/cell/h) with saturable kinetics characterized by Michaelis-Menten constant $K_{m,1}$ (genomes/cell) and consumed by self-degradation with a rate constant $k_{d,1}$ (1/h) and

condensation with vDNA to form vDNA-loaded capsids with a rate constant $k_{s,C}$ (cell/genomes/h) (see Fig. 2A). $Tp5_1$ synthesis is potentially regulated by feedback inhibition by $Tp5_1$ (models 4-6) and $Tp5_1$ degradation is potentially modulated following the trend of increasing $vDNA_{tot}$ (models 2-6) (Figs. S2-S7). Cytoplasmic-localized, tegument-associated $Tp5_2$ proteins are synthesized, degraded, and regulated in a similar manner to $Tp5_1$ proteins. Production and consumption of intranuclear loaded capsids and intracytoplasmic viral particles follow from mass balance. Finally, intracellular viral particles are consumed by leaving the infected host cell and entering the surrounding media (k_{ex}; 1/h). In Eq. 9, V_{media} is defined experimentally as 1 ml and V_{cell} is defined as the total cellular volume of 1×10^6 cells in each well approximated at 0.002 ml [66].

During parameter estimation, it was necessary to account for normalization and the contribution of intracellular capsids and/or particles to the measurement of Tp5 proteins since the starting material subjected to immunoblot was an unfractionated, whole-cell lysate. We accounted for the contribution(s) of capsid and particle by, first, constraining the parameter estimation algorithm such that subsequent unnormalized state variables were at least 1 order of magnitude smaller at 96 hpi and $vDNA_{in,0}$ = 131 genomes/cell to provide a thermodynamic driving force toward infectious virus production (e.g., $Tp5_1(96,131)$ = 5000 genomes/cell, while Capsid(96,131) = 500 genomes/cell). We then summed the relevant species in units of genomes per cell and normalized this quantity to the maximum of this sum at t = 96 hpi and $vDNA_{in,0}$ = 131 genomes per cell to obtain a quantity comparable to the experimental data. See SI Methods for further information on model parameterization and comparison.

ACKNOWLEDGEMENTS

The authors are thankful to Drs. Amy Hudson, Said Audi, Nate Ledeboer, and to the Terhune, Dash, Hudson, and Boger Lab members for their input during the development of this work. We thank Drs. Justin Reitsma, Kristen Westdorp, and Andy Greene as well as the Biotechnology and Bioengineering Center at MCW for past creative and innovative discussions underlying these studies. We thank Drs. Eain Murphy, Michael Nevels, and Tom Shenk for providing recombinant HCMV, UL123 plasmid, and antibodies against HCMV proteins, respectively. We also thank Katie Cataldo for constructing the CDKN1A plasmid. Finally, model schematics were generated using BioRender.com.

This work was supported by the National Institutes of Allergy and Infectious Disease Division of the National Institutes of Health under award number R21AI149039 to SST and RKD and 5R01AI083281 to SST. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Zuhair, M., et al., *Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis.* Rev Med Virol, 2019. **29**(3): p. e2034.
2. Nassetta, L., D. Kimberlin, and R. Whitley, *Treatment of congenital cytomegalovirus infection: implications for future therapeutic strategies.* J Antimicrob Chemother, 2009. **63**(5): p. 862-7.
3. Yinon, Y., D. Farine, and M.H. Yudin, *Screening, diagnosis, and management of cytomegalovirus infection in pregnancy.* Obstet Gynecol Surv, 2010. **65**(11): p. 736-43.
4. Styczyński, J., *Prophylaxis vs preemptive therapy in prevention of CMV infection: new insight on prophylactic strategy after allogeneic hematopoietic cell transplantation.* Acta Haematologica Polonica, 2020. **51**(1): p. 17-23.
5. El Helou, G. and R.R. Razonable, *Letermovir for the prevention of cytomegalovirus infection and disease in transplant recipients: an evidence-based review.* Infect Drug Resist, 2019. **12**: p. 1481-1491.
6. Razonable, R.R. and A. Humar, *Cytomegalovirus in solid organ transplant recipients-Guidelines of the American Society of Transplantation Infectious Diseases Community of Practice.* Clin Transplant, 2019. **33**(9): p. e13512.
7. Fields, B.N., D.M. Knipe, and P.M. Howley, *Fields virology*. 3rd ed. ed. 1996, Philadelphia: Lippincott-Raven Publishers. 2447-2525.
8. Goodrum, F., *Human Cytomegalovirus Latency: Approaching the Gordian Knot.* Annu Rev Virol, 2016. **3**(1): p. 333-357.
9. Ryckman, B.J., et al., *Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion.* J Virol, 2006. **80**(2): p. 710-22.
10. Scrivano, L., et al., *HCMV spread and cell tropism are determined by distinct virus populations.* PLoS Pathog, 2011. **7**(1): p. e1001256.
11. Nguyen, C.C. and J.P. Kamil, *Pathogen at the Gates: Human Cytomegalovirus Entry and Cell Tropism.* Viruses, 2018. **10**(12).

12. Chambers, J., et al., *DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression.* J Virol, 1999. **73**(7): p. 5757-66.
13. Weekes, M.P., et al., *Quantitative temporal viromics: an approach to investigate host-pathogen interaction.* Cell, 2014. **157**(6): p. 1460-1472.
14. Stern-Ginossar, N., et al., *Decoding human cytomegalovirus.* Science, 2012. **338**(6110): p. 1088-93.
15. Vriend, G., *WHAT IF: A molecular modeling and drug design program.* Journal of Molecular Graphics, 1990. **8**(1): p. 52-56.
16. Huang, S.M., et al., *The utility of modeling and simulation in drug development and regulatory review.* J Pharm Sci, 2013. **102**(9): p. 2912-23.
17. Huang, S.M. and M. Rowland, *The role of physiologically based pharmacokinetic modeling in regulatory review.* Clin Pharmacol Ther, 2012. **91**(3): p. 542-9.
18. Zhao, P., et al., *Applications of physiologically based pharmacokinetic (PBPK) modeling and simulation during regulatory review.* Clin Pharmacol Ther, 2011. **89**(2): p. 259-67.
19. Terhune, S.S., et al., *Network mechanisms and dysfunction within an integrated computational model of progression through mitosis in the human cell cycle.* PLoS Comput Biol, 2020. **16**(4): p. e1007733.
20. Jung, Y., et al., *Crosstalk between Plk1, p53, cell cycle, and G2/M DNA damage checkpoint regulation in cancer: computational modeling and analysis.* NPJ Syst Biol Appl, 2021. **7**(1): p. 46.
21. Goldbeter, A., *A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase.* Proceedings of the National Academy of Sciences, 1991. **88**(20): p. 9107-9111.
22. Gerard, C. and A. Goldbeter, *Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21643-8.

23. Iwamoto, K., et al., *Mathematical modeling of cell cycle regulation in response to DNA damage: exploring mechanisms of cell-fate determination.* Biosystems, 2011. **103**(3): p. 384-91.
24. Guedj, J., et al., *Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. 3991-6.
25. Kitagawa, K., et al., *A PDE multiscale model of hepatitis C virus infection can be transformed to a system of ODEs.* J Theor Biol, 2018. **448**: p. 80-85.
26. Rong, L., et al., *Analysis of hepatitis C virus decline during treatment with the protease inhibitor danoprevir using a multiscale model.* PLoS Comput Biol, 2013. **9**(3): p. e1002959.
27. Vardi, N., S. Chaturvedi, and L.S. Weinberger, *Feedback-mediated signal conversion promotes viral fitness.* Proc Natl Acad Sci U S A, 2018. **115**(37): p. E8803-E8810.
28. Mayer, B.T., et al., *Transient Oral Human Cytomegalovirus Infections Indicate Inefficient Viral Spread from Very Few Initially Infected Cells.* J Virol, 2017. **91**(12).
29. Rose, J., et al., *Novel decay dynamics revealed for virus-mediated drug activation in cytomegalovirus infection.* PLoS Pathog, 2017. **13**(4): p. e1006299.
30. Chaturvedi, S., et al., *A molecular mechanism for probabilistic bet hedging and its role in viral latency.* Proc Natl Acad Sci U S A, 2020. **117**(29): p. 17240-17248.
31. Vanarsdall, A.L. and D.C. Johnson, *Human cytomegalovirus entry into cells.* Curr Opin Virol, 2012. **2**(1): p. 37-42.
32. Dittmer, A., et al., *Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-D-ribonucleosides.* J Virol, 2005. **79**(23): p. 14660-7.
33. Kregler, O., et al., *Brefeldin A inhibits expression of DNA packaging proteins and nucleocapsid formation of human cytomegalovirus.* FEBS Lett, 2009. **583**(7): p. 1207-14.
34. Guo, P., et al., *Biological Nanomotors with a Revolution, Linear, or Rotation Motion Mechanism.* Microbiol Mol Biol Rev, 2016. **80**(1): p. 161-86.

35. Gentry, B.G., E. Bogner, and J.C. Drach, *Targeting the terminase: An important step forward in the treatment and prophylaxis of human cytomegalovirus infections.* Antiviral Res, 2019. **161**: p. 116-124.
36. Schwartz, C., et al., *The ATPase of the phi29 DNA packaging motor is a member of the hexameric AAA+ superfamily.* Virology, 2013. **443**(1): p. 20-7.
37. Muller, C., et al., *Structures and Divergent Mechanisms in Capsid Maturation and Stabilization Following Genome Packaging of Human Cytomegalovirus and Herpesviruses.* Life (Basel), 2021. **11**(2).
38. Haynes, R.J., et al., *Standard reference material 2366 for measurement of human cytomegalovirus DNA.* J Mol Diagn, 2013. **15**(2): p. 177-85.
39. Schilling, E.M., M. Scherer, and T. Stamminger, *Intrinsic Immune Mechanisms Restricting Human Cytomegalovirus Replication.* Viruses, 2021. **13**(2).
40. Kalser, J., et al., *Differences in Growth Properties among Two Human Cytomegalovirus Glycoprotein O Genotypes.* Front Microbiol, 2017. **8**: p. 1609.
41. Munger, J., et al., *Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy.* Nat Biotechnol, 2008. **26**(10): p. 1179-86.
42. Koyuncu, E., et al., *Saturated very long chain fatty acids are required for the production of infectious human cytomegalovirus progeny.* PLoS Pathog, 2013. **9**(5): p. e1003333.
43. Mokry, R.L., et al., *Nitric Oxide Circumvents Virus-Mediated Metabolic Regulation during Human Cytomegalovirus Infection.* mBio, 2020. **11**(6): p. e02630-20.
44. Motulsky, H.J. and A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression*. 2003, San Diego, CA: GraphPad Software, Inc. 134-151.
45. Mariamé, B., et al., *Real-Time Visualization and Quantification of Human Cytomegalovirus Replication in Living Cells Using the ANCHOR DNA Labeling Technology.* Journal of Virology, 2018. **92**(18): p. e00571-18.
46. Cardone, G., et al., *Procapsid assembly, maturation, nuclear exit: dynamic steps in the production of infectious herpesvirions.* Adv Exp Med Biol, 2012. **726**: p. 423-39.

47. Strang, B.L., et al., *Human cytomegalovirus UL44 concentrates at the periphery of replication compartments, the site of viral DNA synthesis.* J Virol, 2012. **86**(4): p. 2089-95.
48. Marschall, M., et al., *The human cytomegalovirus nuclear egress complex unites multiple functions: Recruitment of effectors, nuclear envelope rearrangement, and docking to nuclear capsids.* Rev Med Virol, 2017. **27**(4).
49. Sadri, S., et al., *Integrated Computational Modeling of the Kinetics and Regulation of NADPH Oxidase 2 Assembly, Activation, Electron Flow, and Superoxide Production.* The FASEB Journal, 2020. **34**(S1): p. 1-1.
50. Zhang, X., et al., *Integrated Computational Model of Lung Tissue Bioenergetics.* Front Physiol, 2019. **10**: p. 191.
51. Zhang, X., et al., *Integrated computational model of the bioenergetics of isolated lung mitochondria.* PLoS One, 2018. **13**(6): p. e0197921.
52. Pillai-Kastoori, L., A.R. Schutz-Geschwender, and J.A. Harford, *A systematic approach to quantitative Western blot analysis.* Anal Biochem, 2020. **593**: p. 113608.
53. Kaspari, M., et al., *Proteasome inhibitor MG132 blocks viral DNA replication and assembly of human cytomegalovirus.* FEBS Lett, 2008. **582**(5): p. 666-72.
54. Winkler, L.L. and R.F. Kalejta, *The 19S proteasome activator promotes human cytomegalovirus immediate early gene expression through proteolytic and nonproteolytic mechanisms.* J Virol, 2014. **88**(20): p. 11782-90.
55. Tran, K., J.A. Mahr, and D.H. Spector, *Proteasome subunits relocalize during human cytomegalovirus infection, and proteasome activity is necessary for efficient viral gene transcription.* J Virol, 2010. **84**(6): p. 3079-93.
56. Fernandes, S.M., et al., *Cytomegalovirus capsid protease: biological substrates are cleaved more efficiently by full-length enzyme (pUL80a) than by the catalytic domain (assemblin).* J Virol, 2011. **85**(7): p. 3526-34.
57. Kalu, N.N., et al., *Nelfinavir inhibits maturation and export of herpes simplex virus 1.* J Virol, 2014. **88**(10): p. 5455-61.

58. Stevenson, E.V., et al., *HCMV reprogramming of infected monocyte survival and differentiation: a Goldilocks phenomenon.* Viruses, 2014. **6**(2): p. 782-807.
59. Hage, S. and M. Marschall, *'Come together'-The Regulatory Interaction of Herpesviral Nuclear Egress Proteins Comprises Both Essential and Accessory Functions.* Cells, 2022. **11**(11).
60. Tomtishen III, J.P., *Human cytomegalovirus tegument proteins (pp65, pp71, pp150, pp28).* Virology Journal, 2012. **9**(22).
61. Kalejta, R.F. and E.R. Albright, *Expanding the Known Functional Repertoire of the Human Cytomegalovirus pp71 Protein.* Front Cell Infect Microbiol, 2020. **10**: p. 95.
62. Bresnahan, W.A. and T.E. Shenk, *UL82 virion protein activates expression of immediate early viral genes in human cytomegalovirus-infected cells.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14506-11.
63. Baldick, C.J., Jr., et al., *Human cytomegalovirus tegument protein pp71 (ppUL82) enhances the infectivity of viral DNA and accelerates the infectious cycle.* J Virol, 1997. **71**(6): p. 4400-8.
64. Rand, U., et al., *A Novel Triple-Fluorescent HCMV Strain Reveals Gene Expression Dynamics and Anti-Herpesviral Drug Mechanisms.* Front Cell Infect Microbiol, 2020. **10**: p. 536150.
65. Glass, D.S., X. Jin, and I.H. Riedel-Kruse, *Nonlinear delay differential equations and their application to modeling biological network motifs.* Nat Commun, 2021. **12**(1): p. 1788.
66. Milo, R., et al., *BioNumbers--the database of key numbers in molecular and cell biology.* Nucleic Acids Res, 2010. **38**(Database issue): p. D750-3.

FIGURE LEGENDS

Fig. 1. Empirical model of HCMV DNA replication predicts saturation kinetics at high MOIs. **(A)** Schematic of an empirical model of HCMV viral genome (vDNA) synthesis during replication. $vDNA_{in}$ represents cell-associated genomes upon infection with decay rate, k_d. $vDNA_{rep}$ represents replicated genomes, and $vDNA_{tot}$ is total contribution of all subspecies. **(B)** Total DNA was isolated from growth-arrested MRC-5 fibroblasts infected at an MOI of 0.1, 0.5, or 5 infectious units per cell (IU/cell) based on 1 x 10^6 cells using HCMV strain TB40/E encoding IE2-(T2A)-eGFP and pp28-mCherry (IE2-2A-eGFP UL99-mCh). Absolute viral (UL123 gene) and cellular (CDKN1A gene) DNA levels were determined between 2-96 hpi. Mean $\pm$ standard deviation (SD) is plotted from three biological replicates and two technical replicates. **(C)** Solid line represents linear regression and closed circles represent data points, correlating MOI (IU/cell) and vDNA (genomes/cell) and shading represents the 95% confidence interval. **(D)** Fit of empirical model (solid curves) to experimental data of $vDNA_{tot}$ (genomes/cell). **(E)** Each curve represents predicted $vDNA_{tot}$ kinetics at a specific $vDNA_{in,0}$. Magenta and cyan curves described in (F). **(F)** Predictive simulations of $vDNA_{tot}$ (solid curves) compared to data collected *post hoc* (closed circles) at $vDNA_{in,0}$ of 6 (Magenta; MOI 0.23) and 0.2 genomes/cell (Cyan; MOI 0.01). Error bars represent SD of three biological replicates and three technical replicates. **(G)** Coloring of predictions represents the fold change of vDNA between 2 and 96 hpi ($vDNA_{tot}$ at 96 hpi/$vDNA_{in,0}$). Closed circles represent $vDNA_{tot}$ data from (D) and (F). **(H)** Model predictions of $vDNA_{tot}$ separated into $vDNA_{in}$ and $vDNA_{rep}$ kinetics at varying $vDNA_{in,0}$.

Fig. 2. Quantitative measurements of state variables of HCMV replication at multiple MOIs. **(A)** Schematic of a framework for HCMV replication starting from $vDNA_{in,0}$ (genomes/cell) to infectious extracellular virus (IU/ml) involving total genomes $vDNA_{tot}$, nuclear Tp5 proteins $Tp5_1$, cytoplasmic proteins $Tp5_2$, capsids C, particles P, rates of synthesis ($k_{s,n}$) and degradation ($k_{d,n}$) for each species n listed. k_{ex} represents the rate of virus release. Additional putative mechanisms (dashed arrows) are postulated for improving model fit to experimental data. **(B)** Immunoblots of a representative nuclear protein, pUL44 ($Tp5_1$) during infection by HCMV TB40/E (IE2-2A-eGFP

UL99-mCh) at $vDNA_{in,0}$ of 3 (MOI 0.1), 14 (MOI 0.5), and 131 (MOI 5) genomes/cell. Whole-cell lysates from infected MRC-5 fibroblasts were collected and analyzed using an antibody against pUL44 ($Tp5_1$). A 96 hpi protein standard from whole-cell lysate infected at an MOI 5 IU/cell with TB40/E-eGFP. Representative total protein is shown. Asterisks (*) indicate bands quantified; data represent two biological replicates for each $vDNA_{in,0}$. **(C)** Total protein (squares) lane volumes normalized to the undiluted standard (Std 1). Standard pUL44 (triangles) band volumes normalized to undiluted Std 1 band volume. Mean ± SD are plotted from two biological replicates totaling six data points for each dilution. **(D)** The pUL44 ($Tp5_1$) band volumes were normalized to total protein to account for loading error. The values were set relative to Std 1 to normalize between membranes and set to a maximum value within the replicate to obtain relative values between 0 and 1. Mean ± SD are plotted from two biological replicates. Quantities of pUL44 from Weekes et al. [13] are shown. **(E)** Immunoblots of a representative cytoplasmic Tp5 protein, pp28 ($Tp5_2$) as in (B). Asterisks (*) indicates pp28-mCherry. The same protein standard from Fig 2B was used. **(F)** Total protein and pp28 signal of the protein standard as in (C). **(G)** The pp28-mCherry ($Tp5_2$) band volumes were normalized to total protein to account for loading error, values were set relative to Std 1, then set to a maximum value within the replicate to obtain values between 0 and 1. Quantities on pp28 from Weekes et al. [13] are shown. **(H)** Titers were determined by infectious units assay and data are the mean ± SD from two biological replicates for each $vDNA_{in,0}$. **(I)** Infectivity was determined by setting UL123 gene copies relative to infectious units from Fig. 2H. Dotted line indicates infectivity of the viral stock inoculum with mean ± SD from two biological replicates.

Fig. 3. Best fit deterministic model describing expression of HCMV temporal profile 5 proteins from viral DNA to extracellular infectious virus. (A) Best fit model predicting increasing Tp5 degradation with increasing vDNA. **(B)** Fit of deterministic model (curves) to the immunoblot data of a representative nuclear $Tp5_1$ protein (pUL44; closed markers) at $vDNA_{in,0}$ of 3, 14, and 131 genomes/cell. Best fit model parameters were estimated using a pseudo-Monte Carlo minimization procedure. **(C)** Two-(left) and three-(right) dimensional model predictions of

802 $Tp5_1$ kinetics at varying $vDNA_{in,0}$. **(D)** Fit of deterministic model (solid curves) to experimental
803 immunoblot data of a representative cytosolic $Tp5_2$ protein (pp28; closed markers). Best fit model
804 parameters were estimated by using pseudo-Monte Carlo minimization procedure. **(E)** Two-(left)
805 and three-(right) dimensional model predictions of $Tp5_2$ kinetics as in (C). **(F)** Fit of deterministic
806 model (solid curves) to normalized experimental viral titer data (closed markers). Viral titers were
807 normalized to the maximum value in each replicate resulting in arbitrary units (AU) to ensure
808 comparable ranges between fitted data sets. Normalized experimental data shown relative to
809 maximum in the data set (MOI 5, 96 hpi). Parameters were estimated using a pseudo-Monte
810 Carlo minimization protocol. **(G)** Two-(left) and three-(right) dimensional model predictions of
811 normalized infectious virus production varying by $vDNA_{in,0}$ and time starting at 24 hpi.
812
813 **Fig. 4. Simulations of viral output show saturation kinetics mirroring experimental**
814 **evidence and predictive of optimal replication from $vDNA_{in,0}$. (A)** The normalized simulations
815 from Fig. 3G were converted back to measurable units of IU/ml using a conversion factor of 10^7
816 IU/ml. Color differences represent the calculated fold change occurring between 24 and 96 hpi.
817 **(B)** Plots of $vDNA_{tot}$ (left) or Virus (right) fold changes versus $vDNA_{in,0}$. Maximal fold change
818 occurs for $vDNA_{tot}$ when $9 < vDNA_{in,0} < 13$ and for Virus when $5 < vDNA_{in,0} < 9$. **(C)** 4D
819 visualization of the relationships between time, $vDNA_{tot}$, vDNA fold change (color), and $Tp5_{1,tot}$
820 ,**(D)** $Tp5_{2,tot}$,and **(E)** relative viral titers. **(F)** Increased data resolution obtained using live-cell
821 imaging during HCMV infection. MRC-5 fibroblasts were infected as described in Fig. 2. Images
822 were captured every 2 hpi using phase contrast, green (460 nm, IE2-T2A-eGFP), and red (585
823 nm, pp28-mCherry) channels. Representative image for different $vDNA_{in,0}$ at 24 and 72 hpi are
824 shown with video in Supplemental Information. **(G)** Single cell measurements were completed at
825 vDNA = 3 showing pp28-mCherry intensities per eGFP-positive area over time starting at
826 approximately 36 hpi. Open triangles represent cells lysing prior to 96 hpi. **(H)** HCMV pp28-
827 mCherry fluorescence signal per well for all inputs in Fig 4F relative to 96 hpi at vDNA of 131
828 genomes/cell and **(I)** predicted $Tp5_2$ kinetics over a replication cycle shows agreement with high-
829 temporal resolution data.

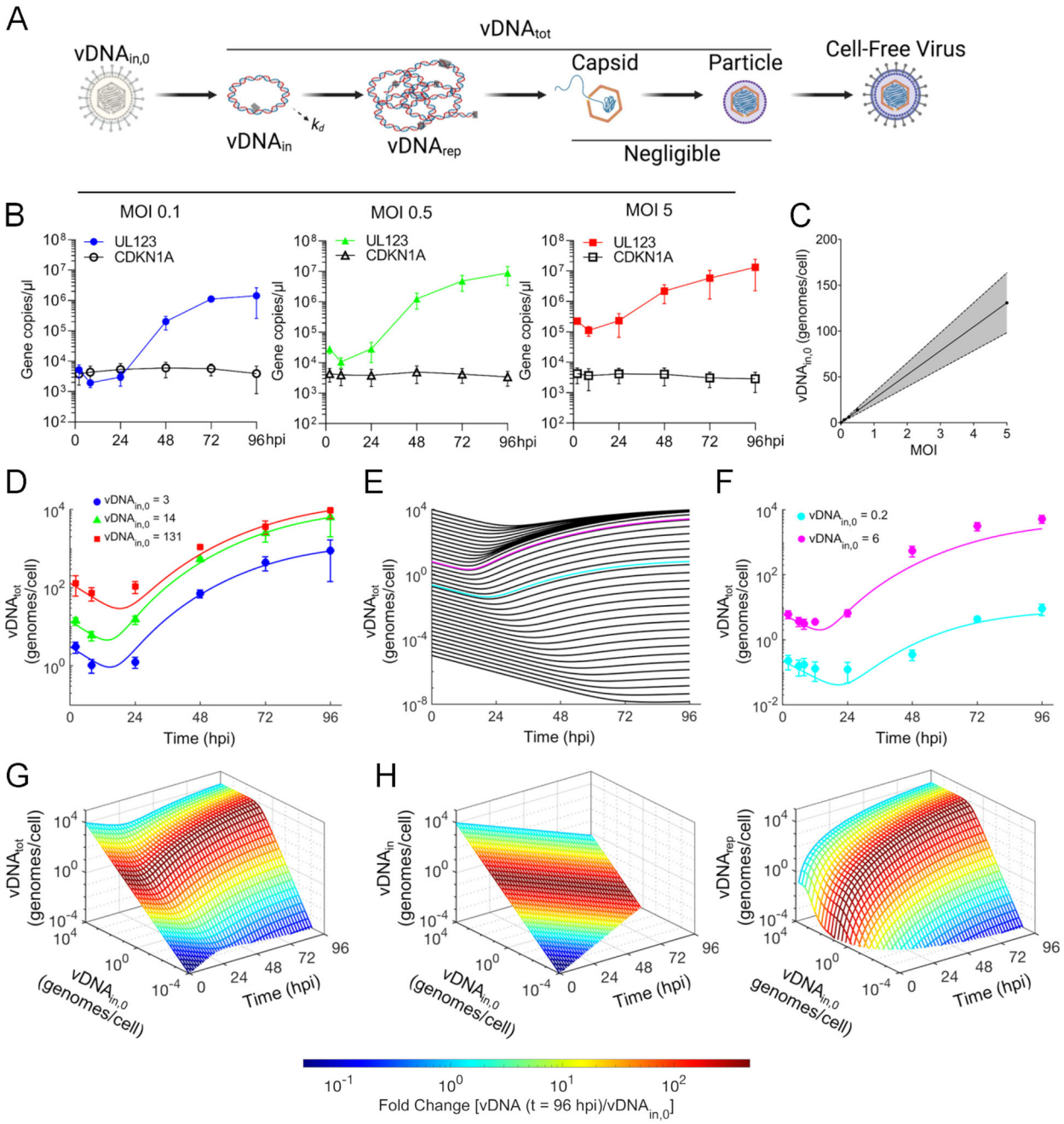
A
vDNA_in,0
vDNA_tot
vDNA_in
k_d
vDNA_rep
Capsid
Particle
Cell-Free Virus
Negligible
B
MOI 0.1
MOI 0.5
MOI 5
UL123
CDKN1A
Gene copies/µl
96hpi
C
vDNA_in,0 (genomes/cell)
MOI
D
vDNA_in,0 = 3
vDNA_in,0 = 14
vDNA_in,0 = 131
vDNA_tot (genomes/cell)
Time (hpi)
E
F
vDNA_in,0 = 0.2
vDNA_in,0 = 6
G
vDNA_in,0 (genomes/cell)
H
vDNA_in (genomes/cell)
vDNA_rep (genomes/cell)
Fold Change [vDNA (t = 96 hpi)/vDNA_in,0]

Figure. 1

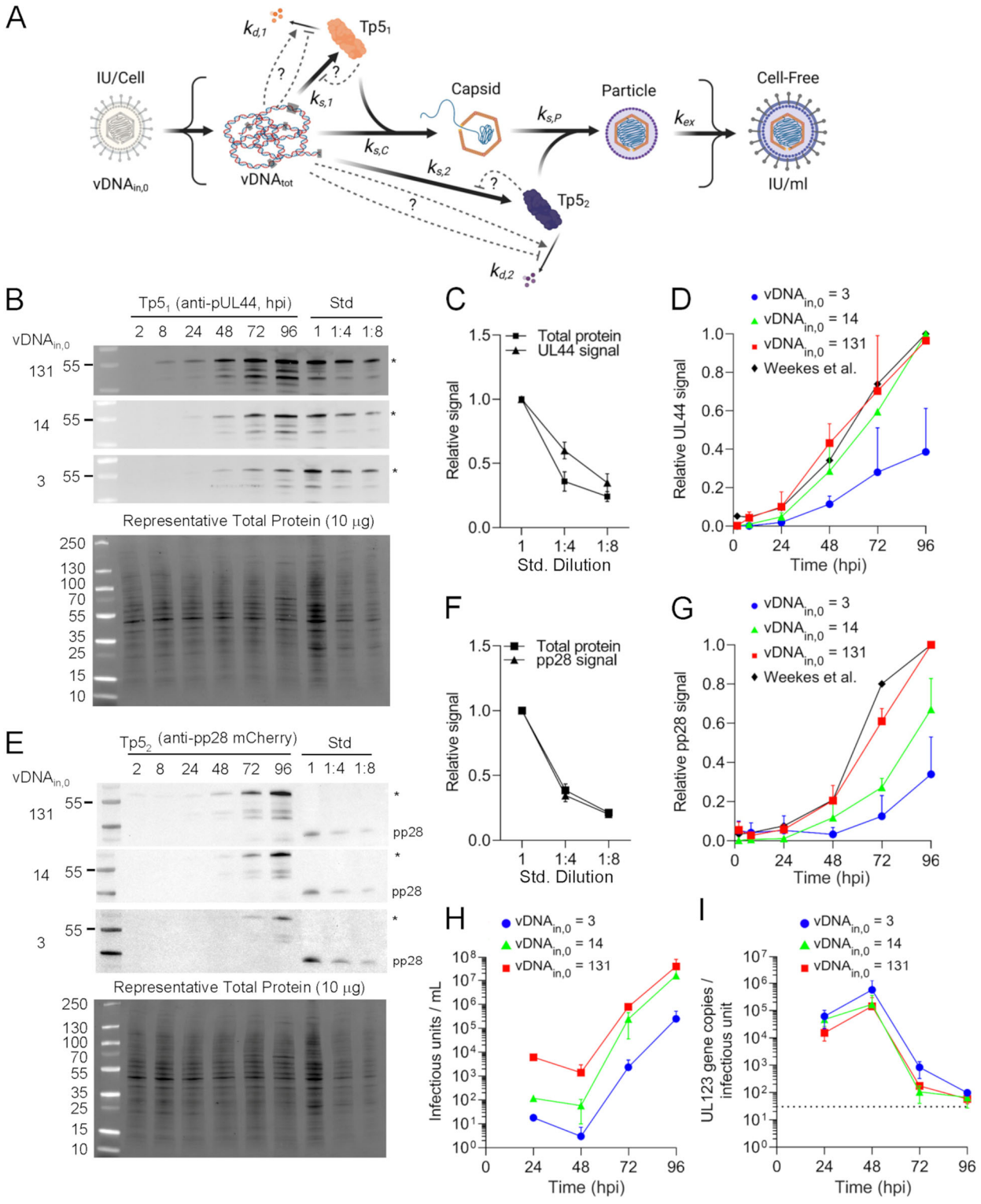
A
IU/Cell
vDNA$_{in,0}$
vDNA$_{tot}$
$k_{d,1}$
Tp5$_1$
$k_{s,1}$
$k_{s,C}$
Capsid
$k_{s,P}$
Particle
k_{ex}
Cell-Free
IU/ml
$k_{s,2}$
Tp5$_2$
$k_{d,2}$
B
Tp5$_1$ (anti-pUL44, hpi)
Std
2 8 24 48 72 96 1 1:4 1:8
vDNA$_{in,0}$
131
14
3
55
Representative Total Protein (10 µg)
250 130 100 70 55 35 25 15 10
C
Total protein
UL44 signal
Relative signal
Std. Dilution
D
vDNA$_{in,0}$ = 3
vDNA$_{in,0}$ = 14
vDNA$_{in,0}$ = 131
Weekes et al.
Relative UL44 signal
Time (hpi)
E
Tp5$_2$ (anti-pp28 mCherry)
pp28
F
Total protein
pp28 signal
G
Relative pp28 signal
H
Infectious units / mL
I
UL123 gene copies / infectious unit

Figure 2

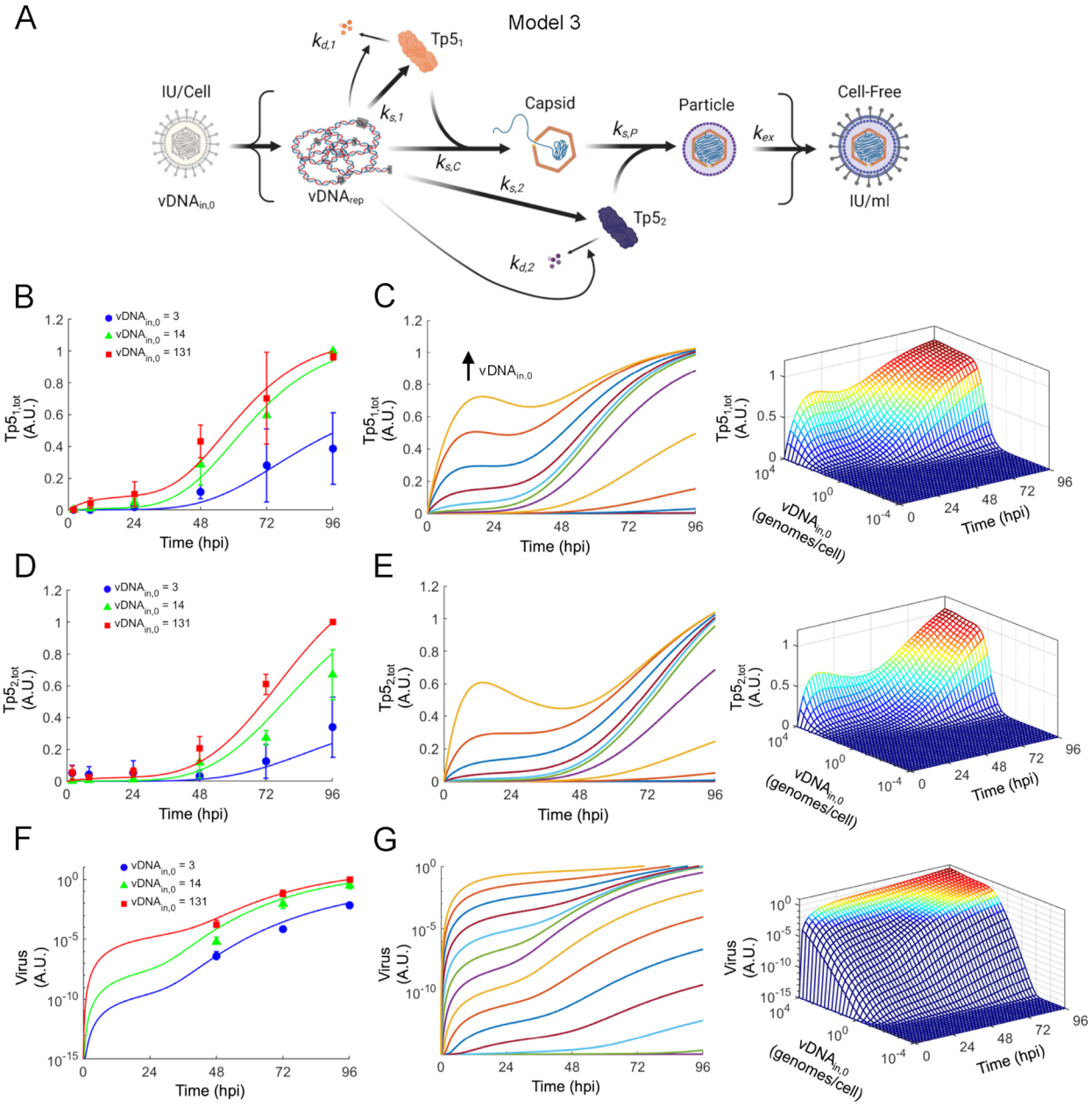
A
Model 3
IU/Cell
vDNA$_{in,0}$
vDNA$_{rep}$
$k_{d,1}$
Tp5$_1$
$k_{s,1}$
$k_{s,C}$
Capsid
$k_{s,P}$
Particle
k_{ex}
Cell-Free
IU/ml
$k_{s,2}$
Tp5$_2$
$k_{d,2}$
B
C
D
E
F
G
vDNA$_{in,0}$ = 3
vDNA$_{in,0}$ = 14
vDNA$_{in,0}$ = 131
Tp5$_{1,tot}$ (A.U.)
Tp5$_{2,tot}$ (A.U.)
Virus (A.U.)
Time (hpi)
vDNA$_{in,0}$ (genomes/cell)

Figure 3

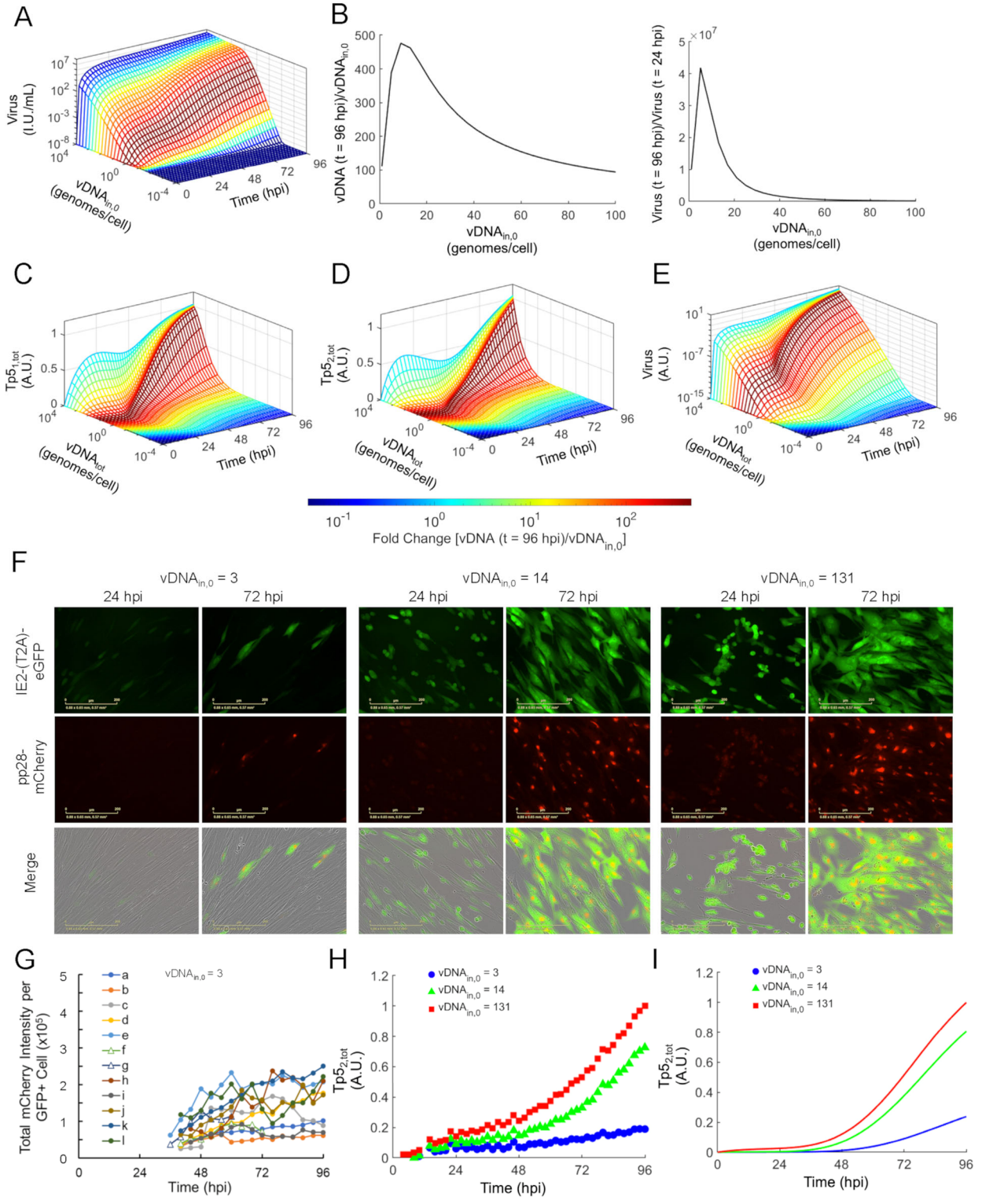
A
B
C
D
E
F
G
H
I
Virus (I.U./mL)
vDNA (t = 96 hpi)/vDNA$_{in,0}$
Virus (t = 96 hpi)/Virus (t = 24 hpi)
vDNA$_{in,0}$ (genomes/cell)
Time (hpi)
Tp5$_{1,tot}$ (A.U.)
Tp5$_{2,tot}$ (A.U.)
vDNA$_{tot}$ (genomes/cell)
Fold Change [vDNA (t = 96 hpi)/vDNA$_{in,0}$]
vDNA$_{in,0}$ = 3
vDNA$_{in,0}$ = 14
vDNA$_{in,0}$ = 131
24 hpi
72 hpi
IE2-(T2A)-eGFP
pp28-mCherry
Merge
Total mCherry Intensity per GFP+ Cell (x10^5)
a
b
c
d
e
f
g
h
i
j
k
l

Figure 4