Evolution of a single-cell predictive model for packaging and budding of viruses based on TEM based measurements

Abha Saxena[1], Suryateja Ravutla[1], Kishalay Mitra[1], Jayanta Chakraborty[2], David W. Murhammer[3], Lopamudra Giri[1]*

Affiliations:

[1]Department of Chemical Engineering, Indian Institute of Technology, Hyderabad, India

[2]Department of Chemical Engineering, Indian Institute of Technology, Kharagpur, India

[3]Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, USA

*Corresponding author. Email: giril@iith.ac.in

Abstract: Although detailed experimental investigations would provide insight into viral infections and vaccine production, building a computational framework is necessary to identify the parameters that regulate the budding and packaging of nucleocapsids. This study shows that a predictive model for the complete infection cycle can be built using nonlinear coupled ODEs and parameter estimation using a Genetic algorithm. Specifically, we have used a dataset containing the occluded virus information, budded virus in infected cells obtained by transmission electron microscopy (TEM). A novel parameter estimation strategy is proposed based on the k-medoid clustering of infected cells. Firstly, we show that the parameter estimation framework can be used for model evolution and selection of the feedback structure. Secondly, we show that the model was capable of capturing the distribution of packaged and unpackaged nucleocapsids in the nucleus, cytoplasm, and plasma membrane, the number of packaged and unpackaged ODV, and polyhedra in the nucleus. The proposed framework assumes importance in generating data for achieving quality by design in the optimization of vaccine/recombinant protein yield.

INTRODUCTION

Viruses are a significant warning to animal health, and many investigations have been performed over the last few decades in understanding the interaction between virus and host cell. Despite this, significant challenges remain in understanding the pattern of protein expression, viral packaging in early, late, and very late infection processes. Large-scale vaccine production and data generation using design remains significantly challenging since viruses undergo a significant amount of genetic mutations during passaging of the viruses (Lopamudra Giri et al. 2012). In this paper, our main objective is to build a computational method that can be used to select a model for a baculovirus infection process using a training dataset, estimate model parameters, and check the model accuracy using a testing dataset. Importantly, the dataset used for constructing the models was generated by single-cell imaging experiments using a transmission electron microscope (TEM) and confocal microscope. Such model building strategy can facilitate the evolution of models that can capture the experimental outcomes of the infection process and the role of protein in controlling the virus assembly from existing datasets generated through imaging of infected cells.

In the last two decades, there has been significant success in developing potent, safe, and affordable vaccines using cell-based production systems (Aubrit et al. 2015; Gallo-Ramírez et al. 2015; Shan et al. 2017; Volz and Sutter 2017). *Autographa californica nucleopolyhedrovirus* (AcMNPV), the most widely studied baculovirus, has been used to produce vaccines (Saxena et al. 2018; Cox 2012; Legastelois et al. 2017) (Cervarix for cervical cancer, Provenge for prostate cancer, Flubok for influenza, Porcilis PCV, and CircoFlex for procine circovirus, Porcilis Pesti for swine fever) and other recombinant proteins (Zitzmann et al. 2017; van Oers, Pijlman, and Vlak 2015). Furthermore, AcMNPV and other baculoviruses are being evaluated for their potential use as biopesticides (Buerger, Hauxwell, and Murray 2007; F. Moscardi 1999; Flavio Moscardi et al. 2011). Virus packaging results in various phenotypes and is regulated by viral protein expression within infected cells. The resulting packaging mechanism has a profound effect in controlling the virus yield (L. Giri et al. 2010).

Baculoviruses have two genetically identical but morphologically distinct forms produced at different times post-infection, i.e., the budded virus (BV) and the occlusion derived virus (ODV). The BV "buds" from the cell as individual virus particles enveloped with Gp64 which is expressed during the early and late phase of the virus infection process (~8 to 48 h post-infection (pi)). Vp39 protein (PVP) is known to be a late protein which is involved in the transport of nucleocapsids to form a budded virus from the nucleus to the plasma membrane. The ODVs are packaged inside polyhedra (King et al. 2011) that accumulate in the cell nucleus later in the infection process (beginning at ~24 h pi and continuing until cell death). ODVs are formed by the envelopment of single or multiple nucleocapsids in microvesicles incorporated with ODV-E66 and the polyhedrin protein (very late) is required for polyhedra formation and a normal occlusion process. Although there are various attempts to explain these processes through molecular biology and cell biology-based experiments, there is a limited investigation for building a mathematical model of the complete infection process.

A mathematical model of viral dynamics will help in understanding the detailed mechanistic insights of viral protein expression and cell-virus interaction to improve vaccine, recombinant protein, or biopesticide production yield. In addition, a method for estimation of model parameters would provide an important tool to predict the viral dynamics, thereby being useful in predicting the protein expression, nucleocapsid formation. This can be further beneficial in developing vaccines.

Various models explaining the viral growth kinetics in bioreactors have been developed (Licari and Bailey 1992; J. Power et al. 1992; De Gooijer et al. 1992; J. F. Power et al. 1994; Kumar and Shuler 1995; J. F. Power and Nielsen 1996; Jang et al. 2000; Haas and Nielsen 2005). These models do not explain the kinetics of virus replication. Mathematical models to study the effect of a multiplicity of infection (MOI) (Zhang and Merchuk 2004; Enden, Zhang, and Merchuk 2005) have also been reported in the literature. Still, these models do not address the intracellular kinetics of baculovirus infection in insect cells. Additionally, defective interfering particle (DIPs) formation has also been modeled (Kirkwood and Bangham 1994; Zwart et al. 2013), but the distribution of nucleocapsids present in a cell population was not considered in the model. Production of virus-like particles (VLPs) in BEVS has also

been modeled (Hu and Bentley 2000; Roldão et al. 2007). Roldao et al.(Roldão et al. 2007) presented a model for the production of Rotavirus-like particles in baculovirus-infected insect cells, focusing mainly on viral DNA replication in the nucleus late mRNA transcription protein synthesis. However, the model does not address the hierarchy of early and very late mRNA transcription and corresponding protein synthesis. Other models characterize the change in infected-cell diameter distribution (Gotoh, Fukuhara, and Kikuchi 2008), the therapeutic effects of antiviral compounds (Orihara et al. 2008), and the cell cycle distribution (Lindeberger et al. 2012). But none of them describe the budded virus formation and packaging of occluded virions. Although structured models have been developed for intracellular baculovirus infection in insect cells, it considers only virus entry through the plasma membrane and transport of the viral genome to the nucleus (Dee and Shuler 1997). Another model was proposed to characterize mRNA expression and key metabolites in the growth medium, cytoplasm, and mitochondria but does not consider the formation of various viral proteins and packaging of occluded and non-occluded virus (Jang et al. 2000). A generic stochastic and deterministic model for intracellular viral kinetics has also been reported (Srivastava et al. 2002). This model accounts for the synthesis and depletion of viral nucleic acids and structural proteins. However, these models do not account for the synthesis of early, late protein or packaging of an occluded and non-occluded virus. Also, the existing models do not elucidate the feedback structures present in the signaling network. Another vital gap present in the literature is that none of these models include a detailed method for kinetic parameter estimation for fitting a system of ODEs to the experimental data obtained by high-resolution microscopy, TEM microscopy. Other ordinary differential equation (ODE) models have been proposed for eukaryotic systems (Karr et al. 2012; Sidorenko and Reichl 2004; Reddy and Yin 1999; Hao, Crouser, and Friedman 2014; Xue, Friedman, and Sen 2009) to describe cell states (Yao, Pilko, and Wollman 2016) through kinetic parameter variations (Ciupe et al. 2006) that capture cell-to-cell variability (Hernández Rodríguez et al. 2019).

In this context, the current study develops a structured mathematical model describing the complete baculovirus infection cycle in insect cells that quantifies packaged and unpackaged virus and expression of viral proteins within different cellular compartments. The proposed model also incorporates feedback

regulation by the Fp25k protein and utilizes a genetic algorithm to numerically estimate kinetic parameters to address cell-to-cell variability in viral infection. Since the proposed model has a number of parameters and the nonlinearity of the ODEs describing the system pose serious challenges in fitting the parameters to experimental data, we implemented GA to find the possible global optima. For these systems, the derivative based classical optimization techniques suffer from convergence to the local optima (Singh et al. 2020). The result shows that the parameters estimated based on a dataset obtained from TEM imaging of infected cells can be used for prediction of protein expression and virus production. Finally, the model was validated using single-cell virus data measured by TEM microscopy. This model provides a tool that enhances the understanding of the intracellular kinetics of virus replication, and the role of various proteins involved in virus packaging and viral transport. Also, the result shows that the incorporation of feedback loop yields higher expression of very late protein that is required for viral occlusion. One important feature of this model is that it can be used to visualize the cell-to-cell variability in viral packaging and protein expression. Thus, the proposed framework can be further utilized to optimize recombinant protein, vaccine, and biopesticide production.

MATERIALS AND METHODS

Data generation

The WT and ST virus-infected cells from passage 5 and 10 were collected at 72 hours post-infection (hpi). Samples collected at 72 hpi were examined using a transmission electron microscope, and single-cell images were captured (Figure 1A). The TEM images were used to count the number of unpackaged nucleocapsids in the nucleus ($VN_{,N}$), nucleocapsids meant to become ODV (VN_{VO}), nucleocapsids meant to become BV (VN_{VB}), nucleocapsids at the plasma membrane ($VN_{,M}$), total ODVs in the cell (VO), ODVs in the nucleus ($VO_{,N}$), ODVs packaged inside polyhedra (VO_{POL}) and polyhedra per cell (POL). The viral infection of the cells was performed as described in Giri et al.(L. Giri et al. 2010).

Mathematical model

The mathematical model is based on the signaling network shown in Figure 1B-C. The single-cell model for the complete baculovirus infection process is represented by coupled ordinary differential equations (ODEs) with various kinetic parameters (see supplementary information). The model includes transient DNA and protein synthesis and feedback regulation by the Fp25k protein. We have used the same model structure for predicting the infection pattern for genetically different viral strains. A detailed description of the complete mathematical model is given in supplementary information. A list of model variables and kinetic parameters is given in Tables 1 and 2, respectively.

Multiple model formulation and model selection

A set of four models were formulated based on a various hypothesis with respect to presence and absence of feedback regulation by the Fp25k protein. Model 1(M1) contains no feedback regulation by the Fp25k protein (Figure 2A). Model 2 (M2) is marked by the presence of only negative feedback regulation of transcription of the Gp64 protein by the Fp25k protein (Figure 2B). Model 3(M3) consists of two positive feedback regulations by the Fp25k protein (Figure 2C). These positive feedback regulations include regulation of ODV-E66 protein transport from cytoplasm to nucleus and regulation of polyhedrin protein transcription. Model 4 (M4) is a combination of M2 and M3 that incorporates both negative and positive feedback regulation by the Fp25k protein as described in M2 and M3 (Figure 2D). The model that fit the experimental data with the minimum possible root mean square error value (RMSE) was considered to be the best model.

Clustering of cells

Cells were clustered based on six features, including the number of nucleocapsids meant to become ODV per cell, nucleocapsids in the nucleus per cell, number of ODVs in the nucleus, and number of ODVs packaged in polyhedra per cell, and the total number of ODVs and polyhedra per cell. The k-medoids clustering (Fu et al. 2020) algorithm was used to cluster cells using the Euclidean distance matrix for the distance calculation. The number of clusters was chosen as nine as k=9 gives rise to minimum Davies-Bouldin index (Society for Industrial and Applied Mathematics 2002). Since the data is six-dimensional,

we performed principal component analysis (PCA) (Jolliffe 1986) to visualize the data in a 2D plane (Figure 3A). PCA was implemented for data representation after clustering of the cells.

Description of model parameters

The proposed model has 33 unknown parameters, out of which 31 parameters were fixed for all the viral-infected cells, whereas DNA replication rate (kr_{DNA}) and RNA transcription (ktr_{RNA}) rate were chosen as free parameters that account for the heterogeneity in viral packaging, polyhedral packaging as well as protein expression.

In order to address the cell-to-cell variability in viral infection, we show the clustering of the infected cell population into nine groups, as shown in Figure 4 (circle A) and Figure 3A. Since the mutations may lead to a change in viral genome length in WT and stabilized viruses, it is possible that the protein expression is significantly altered due to a change in codons coding for amino acids (Coulon et al. 2014). Hence it is assumed that the infection process for a mixed virus population involved a variation in the rate of RNA synthesis or transcription (ktr_{RNA}) and was chosen as free parameters to model variability across clusters (Figure 4, circle B). It was assumed that every cluster has a different rate of RNA synthesis, but all the cells within a cluster have the same rate of RNA synthesis. Therefore, the model assumes nine RNA synthesis rates $(ktr_{RNA}1 - ktr_{RNA}9)$ corresponding to 9 clusters.

Although a fixed MOI is assumed for the infection of each cell, it is possible that all the cells are not infected with the same number and type of virus (Saxena et al. 2020). Also, since viral passaging yields different types of mutations, it is assumed that there is a variation in the rate of DNA replication rate (kr_{DNA}) in infected cells (Lopamudra Giri et al. 2012).

In order to address the variation within the cluster, we choose the rate of viral DNA replication (k_{32}) as a free parameter to model the variability in a cluster (Figure 4, circle C). Here, it is assumed that each of the infected cells within and across the cluster will have a different rate of DNA replication resulting in 80 DNA replication rates $(kr_{DNA}1 - kr_{DNA}80)$ corresponding to 80 cells. The model parameters were estimated using six data sets (training dataset), and the remaining two data sets (testing dataset) were used

for model validation. The training data sets used for parameter estimation using GA consists of $VN_{,N}$, VN_{VO}, VO, $VO_{,N}$, VO_{POL}, and POL. The testing data sets used for the validation of the proposed model consists of VN_{VB} and $VN_{,M}$. The kinetic parameters used to address the cell-to-cell variability in virus packaging and viral protein expression were chosen based on previous literature regarding Fp25k protein feedback regulation (Li et al. 2015) and parametric sensitivity analysis.

Parameter estimation

Specifically, GA was implemented to perform the parameter estimation since it provides effective search procedures in various applications, especially those involving nonlinearity (Mirjalili 2018)_The objective function was defined as the root mean squared error (RMSE) between the experimental measurement and the model output.

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \begin{matrix} \left(Exp_k - Sim_k\right)^2_{VN_{,N}} + \left(Exp_k - Sim_k\right)^2_{VN_{VO}} + \left(Exp_k - Sim_k\right)^2_{VO_{POL}} \\ + \left(Exp_k - Sim_k\right)^2_{VO_{,N}} + \left(Exp_k - Sim_k\right)^2_{VO} + \left(Exp_k - Sim_k\right)^2_{POL} \end{matrix}}$$

$$n: number\ of\ cells; k: cell\ index; \exp: Experimental\ data; \sim: Simulated\ data$$

In each iteration, a new population set of decision variables (fixed and free parameters) containing the probable solution set was computed, resulting in the new generation. In each generation selection, crossover and mutation were performed (crossover fraction = 0.8, population size=200, and fitness limit = -inf). New generations were produced until the desired tolerance limit on the error value was reached. The set of parameters estimated through GA with the lowest RMSE (Saxena et al. 2019) were selected. Different combinations of fixed and free parameters were tested based on the results of parametric sensitivity analysis. The difference between the simulation and experiment was aggregated, and a RMSE was calculated to assess the goodness of fit for each combination of fixed and free parameters (Table S1).

Experimental data generation for design-based optimization

The proposed model was used to generate the data on infection characteristics for 10^5 cells cultured in a shaker flask. The fixed parameters estimated using GA were used for simulation as provided in Table 2. The values of free parameter (kr_{DNA}) for 10^5 cells were generated using Monte-Carlo simulation

following Birnbaum Saunders distribution (Saxena et al. 2020). The data on the amount of budded virus, polyhedral and protein levels in infected cells were generated at various MOIs (2,4,6,8,10,12) and TOHs (12,24,36,48,60,72 hours post infection) using factorial design.

Design of Experiment (DOE) based studies for data generation to optimize infection process

Two level factorial design was employed to determine the optimal condition for virus production using independent variables as multiplicity of infection (MOI) and time of harvest (TOH). The data obtained from simulations was fitted using multiple regression to a second-order polynomial equation as:

$$Y_{fit} = \beta_0 + \sum \beta_1 . x_1 + ¿ \sum \beta_2 . x_2 + ¿ \sum \beta_3 . x_1 . x_2 + ¿ \sum \beta_4 . x_1^2 + ¿ \sum \beta_5 . x_2^2 ¿ ¿ ¿ ¿$$

Here, $'x_1^{'}$ represents MOI and $'x_2^{'}$ represents TOH. In order to evaluate the quality, the median value of all responses was used at various MOI and TOH. Response surfaces were generated to analyze the optimal TOH, and MOI required for virus production and protein expression for the particular viral strain under consideration.

RESULTS

Comparison of multiple models and validation of feedback regulation by the Fp25k protein

In order to analyze the role of feedback loops, four possible models were constructed based on different hypotheses for feedback regulation of early, late, and very late proteins via Fp25k protein as shown in Figure 2A-2D. Figure 2A represents the absence of feedback regulation by Fp25k protein (Hypothesis 1). Figure 2B illustrates negative feedback regulation of Gp64 protein transcription by the Fp25k protein (Hypothesis 2). Figure 2C represents the presence of positive feedback regulation of ODV-E66 protein transport from the cytoplasm to the nucleus and polyhedron protein transcription by the Fp25k protein (Hypothesis 3). Figure 2D shows the schematic of hypothesis 4 that includes both negative and positive feedback regulation by the Fp25k protein on Gp64 and ODV-E6 expression.

Next, the dynamics of nucleocapsids (VN), ODV (VO), polyhedral (POL) formation, and protein expression of all four models were investigated. Figure 2E-2J illustrates the time courses for $VN_{,N}$, $VN_{VO,}$ VO, VO_{POL}, and POL formation. These results demonstrate that the negative feedback loop

leads to a decrease in amplitude of $VN_{,N}$, $VN_{VO.}VO$, VO_{POL}, and POL production. On the other hand, the positive feedback loop is able to increase the level of these variables. The feedback regulation in the model was validated with experimental results given by Sholfen Li et al. (Li et al. 2015). Figure 2K and 2L show that the level of the Fp25k protein regulates the protein expression as well as packaged virus including VN_{VO}, $POE_{,N}$, VO, $PPH_{,N}$, POL, PGP, VN_{VB}, $VN_{,N}$, and $POE_{,C}$ (Table 1). The Fp25k protein level was mainly regulated by increasing and decreasing the rate of Fp35k synthesis, ktl_{FP}. The result shows that the increased level of Fp25k protein negatively regulates Gp64 protein transcription and formation of VN_{VB}, $VN_{,N}$, and $POE_{,C}$ (Figure 2K) as found in Sholfen Li et al.(Saxena et al. 2018; Li et al. 2015). Furthermore, the simulation results show that the decreased level of Fp25k protein positively regulates the late and very late protein expression and occluded virus including VN_{VO}, $POE_{,N}$, VO, $PPH_{,N}$, and POL (Figure 2L).

Comparison of virus distribution in cell population from TEM experiments and simulation

Next, we compared the distribution of budded and occluded virus production, polyhedral formation obtained from the experiment (Figure 5A-5D), and simulation results using the parameters estimated for the four models. Specifically, we analyze if the estimated kinetic parameters are able to address cell-to-cell variability in the viral infection process. Figure 5E-5J compares simulation results from all four models with experimental data through violin plot representation corresponding to the training dataset. The result shows that model 3 and model 4 are able to capture the distribution of VN_{VO}, $POE_{,N}$, VO, $PPH_{,N}$, and POLobtained from experiments (Figure 5E-5J). In contrast, the data generated by models 1 and 2 are significantly different from experimental data (Figure 5E-5J). In addition to this, we also performed the goodness of fit analysis. The results show that the experimental and simulated data are similar for models 3 and 4. On the other hand, the simulation results are not able to emulate the pattern obtained from experiments when models 1 and 2 are used (Figure 5K-5P), resulting in rejection of these

models based on hypotheses 1 and 2. Overall, the result shows that the presence of positive feedback is essential in depicting distribution in virus production and packaging pattern.

Model validation using the testing dataset obtained from TEM analysis

Although the proposed model with positive and negative feedback captures the training dataset, we further validate models 3 and 4 with respect to the testing dataset. In order to select the best model, the simulated datasets were compared with the testing datasets for VN_{VB}, and $VN_{,M}$ obtained from TEM experiments (Figure 5Q-5R). Figure 5S shows that there is no significant difference in experimental and simulation results for VN_{VB} for model 4. However, there is a significant difference between the experimentally obtained distribution of $VN_{,M}$ and the simulated distribution of $VN_{,M}$ obtained from model 3 (Figure 5T). Therefore, we reject model 3 based on hypothesis 3.

Although the parameter estimation using GA was not performed based on the testing dataset, model 4 was able to predict the testing dataset using the parameters estimated from the training dataset. Overall, the results indicate that the presence of both positive and negative feedback yields a robust model that can capture the inherent heterogeneity in the infection process.

Model evolution based on RMSE between experimental and simulation data in single cells

Although model 4 was able to predict the distribution of virus particles, polyhedra, and protein expression in a cell population, we further performed RMSE analysis to test the predictability of the models for the data obtained for individual cells. Towards this, we compared all four models based on the root mean squared error (RMSE) between experiment and simulation (Figure 6). The simulation results for model 1 show the highest RMSE for training data V_1 ($VN_{,N}$), V_2 (VN_{VO}), V_3 (VO), V_4 ($VO_{,N}$), V_5 (VO_{POL}), V_6 (POL), and testing data V_7 (VN_{VB}), V_8 ($VN_{,M}$) resulting from error between experimental and simulated results corresponding to each infected cell present in the dataset (Figure 6A). In comparison to model 1, the simulation results of model 2 show a reduction in RMSE for V_1, V_2, V_3, V_4, and V_5, but does not show any improvement in RMSE for polyhedra production (V_6 , Figure 6B). Compared to models 1 and 2, simulation results obtained from model 3 show a further reduction in RMSE for polyhedra production V_6

(Figure 6C). However, model 3 yields higher RMSE for nucleocapsid at the membrane (V_8) present in the testing dataset. The RMSE obtained for all the individual cells from model 4 was found to be at a minimum compared to models 1-3 (Figure 6D). RMSE analysis for individual cells shows that model 4 with both positive and negative feedback loops is essential for prediction of polyhedra production and nucleocapsids present in the membrane.

Parametric sensitivity analysis for choosing the free and fixed kinetic parameter

The above method is based on the choice of the fixed and free parameters from the clustering of the infected cells and sensitivity analysis of the kinetic parameters. In this section, we show the grouping of parameters according to their sensitivity for the proposed model (model 4). Figure 7 shows the effect of changing kinetic parameter (Table 2) values on unpackaged nucleocapsids in the nucleus ($VN_{,N}$), nucleocapsids meant to become ODVs (VN_{VO}), total ODVs in the cell (VO), ODVs in the nucleus ($VO_{,N}$), ODVs packaged inside polyhedra (VP_{POL}) and polyhedra per cell (POL) that were considered as the training dataset. Figure 7A and 7B show the effect of decrease and increase (20%), respectively, in kinetic parameters (k_1-k_{33}) (Table 2) from their basal values for packaged and unpackaged nucleocapsids (V_1-V_6). The kinetic parameters are categorized to have no (0% change), low (<10% change), medium (<15% change), and high (>15% change) sensitivity. The kinetic parameters (k_1-k_6) associated with Gp64 protein formation (translation, transport, and degradation of Gp64 protein) were found to have minimal sensitivity for the budded and occluded virus production considered (Figure 7A and 7B). The result shows that the parameters representing virus internalization from extracellular into the nucleus (k_7-k_{11}) and nucleocapsid formation (k_{12}-k_{13}) were found to have low sensitivity for the production of viruses in various compartments (V_1-V_6).

The parameters responsible for the rate of protein degradation within the nucleus, VP39 protein translation and packaging nucleocapsids, ODVs, and polyhedra (k_{14}-k_{22}), were found to have medium sensitivity for the viral phenotypes. In contrast, the analysis shows that the kinetic parameters (k_{23}-k_{28}) controlling the rate of virus DNA transport to virogenic stroma $(kvt_{N,VS})$, transcription of RNA (ktr_{RNA} ¿

, translation of Fp25k (ktl_{FP}¿, ODV-E66 (ktl_{OE}¿, polyhedrin (ktl_{PH}¿, and vp39 (ktl_{VP}¿ proteins and their transport from the cytoplasm to the nucleus ($kpt_{C,N}$¿affect the responses significantly. Furthermore, the degradation rate of DNA (kd_{DNA}¿and RNA (kd_{RNA}¿ in the nucleus, degradation of protein in the cytoplasm ($kdp_{,C}$), and translation of capsid proteins (ktl_{CP}¿were found to be highly sensitive (k_{29}-k_{31}) for the viral responses considered. Finally, the result shows that the rate of viral DNA replication (kr_{DNA}¿ and degradation of DNA (kd_{DNA}) can be used to regulate all packaging patterns considered in the model (V_1-V_6). However, since the variation in DNA degradation is rather unlikely, the framework was established based on the two free parameters, rate of RNA synthesis (k_{28}:ktr_{RNA}) and the rate of viral DNA replication (k_{32}:kr_{DNA}). The analysis presented here shows that the free parameters chosen for accounting the variation due to genetic modifications and other factors fall under the category of high sensitivity. We also performed a detailed RMSE analysis in which various combinations of free parameters were analyzed. The result shows that the RMSE is minimum when rate of RNA synthesis (k_{28}:ktr_{RNA}) and the rate of viral DNA replication (k_{32}:kr_{DNA}) were taken as the free parameters (Table S1).

Analysis of fixed and free parameters in the infected cell population

One of the major benefits of the proposed scheme of model fitting and model parameterization is to depict the physiological shift of the new strains of viruses during passaging. In order to address this, after fitting a single model structure and parameterization, we performed a detailed comparison among parameters obtained from the cells infected with different viral strains from various passages. Figure 3B shows the box plot representation of the estimated rate of DNA replication corresponding to infected cells within each cluster (Figure 3A) as mentioned above. The remaining parameters of the model were assumed to be fixed parameters and have the same values for all cells within and across the cluster (Figure 6). Figure S1A shows the heatmap representation of the number of nucleocapsids, ODV, and polyhedra at 72 hpi obtained from model 4, along with the estimated rate of DNA replication and RNA synthesis (Figure

S1A). Figure S1B shows the violin plot representation of these parameters corresponding to the wild-type (WT) and stabilized-type (ST) virus. Interestingly, the result shows significant variation in the rate of DNA replication and RNA synthesis for WT and ST virus-infected cells (Figure S1B). Further, we compared the time course of all the model variables for the training dataset across nine clusters (Figure S2). Overall, the results show that the same model structure can be fitted for cells infected with two different viruses (WT and ST), and corresponding parameters can be compared to gain insight into the physiological shift in DNA replication and RNA synthesis.

In order to show the convergence during parameter estimation using GA for model 4, we present the fixed and free parameters at 500, 1200, and 1908 generations (Figure S3). The result shows that the parameter values are not changing much after 1900 generations. The estimated parameter values for the best model (model 4) are given in Table 2. Although the minimum RMSE was obtained for parameters estimated using GA, we also compare the results for parameter estimation using interior point method (fmincon function in MATLAB) and genetic algorithm (Figure S4). The result shows that finding a global solution by providing a proper initial guess in a classical optimizer could be a difficult task for a large-size multi-dimensional optimization formulation for the proposed system of ODEs. In contrast, GA may provide a global solution and a better parameter set as it does not use the derivative information for convergence in the search process.

Application in quality by design

It is challenging to obtain data with factorial design as generation of data on nucleocapsids and other responses for large number of cases is rather expensive (Singh et al. 2020; Puente-Massaguer, Lecina, and Gòdia 2020). In this context, we show a proof of concept that the proposed framework can be used in analyzing the protein level/BV production using design of experiment. Since, the MOI and time-of-harvest (TOH) is known to influences the vaccine/recombinant protein/biopesticide production, we choose to analyze four different responses including protein expression, BV production, polyhedrin protein expression and polyhedral production as a function of MOI and TOH. The interpolation was performed based on data generated at 2, 4, 6, 8, 10, 12 MOI and 2, 24, 36, 48, 60, 72 TOH using a

quadratic function. The result indicates that there is an increase in virus release (BV) and virus assembly (polyhedra) with increase in MOI and TOH. However, the budded virus shows optimal release at 48-60 TOH (Figure 8A) and polyhedral shows optimal release at 60-72 TOH (Figure 8B) at around 12 MOI. The early protein, Gp64 showed the maximal production at 12-18 TOH and 10-12 MOI (Figure 8C). In contrast, the polyhedrin protein (very late protein) produces higher production at low and very high moi and 48-60 TOH (Figure 8D). Finally, we conclude that the proposed structured model can be used for obtaining the data for various designs for a particular virus with its passaging status.

DISCUSSION

In this work, a mathematical model describing the intracellular kinetics of baculovirus infection in different cellular compartments of insect cells is described (Figure 1). The current analysis shows that the cell-to-cell variability can be attributed to the variability in the DNA replication rate. Such variation in DNA replication rate has been reported by Shufen et al.(Li et al. 2015), which shows that the Fp25k protein regulates early and late protein transcription. Since DNA is intimately involved in RNA transcription, any mutation or change in DNA replication rate would result in cell-to-cell variability. Hence, we chose the rate of DNA replication (kr_{DNA}), and RNA transcription (ktr_{RNA}¿ to address variability within and between clusters, respectively. Although we have taken kr_{DNA} and ktr_{RNA} as the free and fixed-parameter, it is possible that other combinations of the free and fixed-parameter may yield a match between experimental and simulated results. In order to rule out other combinations, we performed parameter estimation for various combinations of free and fixed parameters as shown in table S2 and calculated the RMSE corresponding to each case. Also, adding more free parameters may result in overfitting, and the model would be too specific for the data presented. Moreover, we have considered a few proteins involved in the infection process in the current model. But there are other proteins including p10, p6.9, Pp34 (Saxena et al. 2018) that are also involved in virus packaging, and further investigation is needed to decipher the role of those proteins and corresponding feedback loops.

One of the major novelties of the model proposed is that the model is developed based on TEM data that was used to quantify the range of NCs, ODVs, and polyhedra production. However, the number of NCs may be underestimated since the data were obtained using cell slices while performing electron microscopy. The accuracy of these data could be improved through the development of an imaging method for the measurement of nucleocapsids inside ODVs and polyhedral based on 3D imaging of the infected cells.

The time constants were chosen such that the model is able to replicate the time for initiation and termination of DNA replication, transcription of early, late, and very late proteins (Table S2, Figure S5) (Roldão et al. 2007). However, one of the major limitations of this work is that the dynamic pattern could not be matched with single cell experimental data. Nevertheless, such data collection remains challenging since it needs live imaging experiments for a single cell. Also, simultaneous measurement of protein expression and the number of nucleocapsids cannot be performed in the same cell since TEM imaging, and fluorescent immunoassay cannot be performed for the same cell. Another limitation of the proposed model is that it considers only one replication cycle, i.e., the reinfection of a budded virus was incorporated. In future we plan to upgrade the model with reinfection and cell death due to bursting.

The proposed framework contains various modules of infection including virus internalization, DNA replication, transcription, protein transport as well as viral packaging and polyhedra. Although the life cycle of baculoviruses is quite unique compared to other DNA and RNA viruses, some of the modules are common for other types of viruses, and the same framework with parameter estimation scheme can be implemented in building models for other viruses. Specially, in the current pandemic situation, many new approaches towards vaccine development have been attempted, giving rise to DNA and RNA vaccines (Kutzler and Weiner 2008; Puente-Massaguer, Lecina, and Gòdia 2020)-(Rappuoli et al. 2014). Although detailed experimental data on cell-based models and bioreactor data would provide insight into the efficacy of vaccines, the safety level to be maintained by the scientists and industries are stringent for minimizing the risk of viral transmission (Dandri, Volz, and Lütgehetmann 2016). Hence, this work addresses the need of expanding the computational toolboxes that are capable of resolving the constant

challenges posed by new viruses, including Zika virus, Ebola virus, or SARS-COV-2 -coronavirus (Ahmed et al. 2021; Resmawan, Resmawan, and Yahya 2020; Zeb et al. 2020; Badgujar, Badgujar, and Badgujar 2020; Schiffer et al. 2016; Qiu et al. 2012). The mathematical model proposed here has substantial application in production of recombinant vaccine for COVID-19 (Wang 2021) . The SARS-CoV-2 spike protein (S protein) has been expressed in insect cells using BEVS system (Wang 2021; Fujita et al. 2020). The expression of the protein can be modelled using the model as the model considers protein association during the virus formation, packaging, assembly. Although quality by design being the key step for process optimization, data acquisition for pilot scale at various conditions is cumbersome and needs manpower. Although the run time for such simulation for a fermenter is high, the proposed approach may save manpower and cost of performing experiments. Finally, the quality can also be determined accurately since it focuses on getting the data for the population and the cell-to-cell variability.

Supplementary Materials

Mathematical Model

Figure S1. Heat map representation of the training and testing data obtained by TEM and confocal microscopy along with the DNA and RNA replication rate for single cells within a population.

Figure S2. Time course of model variables in single cells within various cluster.

Figure S3. The progression of fixed and free parameters with generations during GA implementation (for the training dataset).

Figure S4. Comparison of parameter estimation results using interior point method (fmincon function in MATLAB) and genetic algorithm.

Figure S5. Comparison of early, late and very late process to validate time function.

Table S1. Design of various combinatorics of free and fixed parameter for which model 4 was tested with respect to experimental data and corresponding RMSEs.

Table S2. Description of time function

References:

1. Ahmed, Idris, Goni Umar Modu, Abdullahi Yusuf, Poom Kumam, and Ibrahim Yusuf. 2021. "A Mathematical Model of Coronavirus Disease (COVID-19) Containing Asymptomatic and Symptomatic Classes." *Results in Physics*. https://doi.org/10.1016/j.rinp.2020.103776.
2. Aubrit, Françoise, Fabien Perugi, Arnaud Léon, Fabienne Guéhenneux, Patrick Champion-Arnaud, Mehdi Lahmar, and Klaus Schwamborn. 2015. "Cell Substrates for the Production of Viral Vaccines." *Vaccine* 33 (44): 5905–12.
3. Badgujar, Kirtikumar C., Vivek C. Badgujar, and Shamkant B. Badgujar. 2020. "Vaccine Development against Coronavirus (2003 to Present): An Overview, Recent Advances, Current Scenario, Opportunities and Challenges." *Diabetes & Metabolic Syndrome* 14 (5): 1361–76.
4. Buerger, Patrick, Caroline Hauxwell, and David Murray. 2007. "Nucleopolyhedrovirus Introduction in Australia." *Virologica Sinica*. https://doi.org/10.1007/s12250-007-0019-y.
5. Ciupe, M. S., B. L. Bivort, D. M. Bortz, and P. W. Nelson. 2006. "Estimating Kinetic Parameters from HIV Primary Infection Data through the Eyes of Three Different Mathematical Models." *Mathematical Biosciences* 200 (1): 1–27.
6. Coulon, Antoine, Matthew L. Ferguson, Valeria de Turris, Murali Palangat, Carson C. Chow, and Daniel R. Larson. 2014. "Kinetic Competition during the Transcription Cycle Results in Stochastic RNA Processing." *eLife* 3 (October). https://doi.org/10.7554/eLife.03939.
7. Cox, Manon M. J. 2012. "Recombinant Protein Vaccines Produced in Insect Cells." *Vaccine*. https://doi.org/10.1016/j.vaccine.2012.01.016.
8. Dandri, Maura, Tassilo Volz, and Marc Lütgehetmann. 2016. "Experimental Models: Cell

Culture and Animal Models." *Molecular and Translational Medicine*. https://doi.org/10.1007/978-3-319-22330-8_2.

9. Dee, K. U., and M. L. Shuler. 1997. "A Mathematical Model of the Trafficking of Acid-Dependent Enveloped Viruses: Application to the Binding, Uptake, and Nuclear Accumulation of Baculovirus." *Biotechnology and Bioengineering* 54 (5): 468–90.
10. De Gooijer, C. D., R. H. Koken, F. L. Van Lier, M. Kool, J. M. Vlak, and J. Tramper. 1992. "A Structured Dynamic Model for the Baculovirus Infection Process in Insect-Cell Reactor Configurations." *Biotechnology and Bioengineering* 40 (4): 537–48.
11. Enden, Giora, You H. Zhang, and José C. Merchuk. 2005. "A Model of the Dynamics of Insect Cell Infection at Low Multiplicity of Infection." *Journal of Theoretical Biology* 237 (3): 257–64.
12. Fujita, Ryosuke, Masato Hino, Takeru Ebihara, Takumi Nagasato, Akitsu Masuda, Jae Man Lee, Tsuguru Fujii, et al. 2020. "Efficient Production of Recombinant SARS-CoV-2 Spike Protein Using the Baculovirus-Silkworm System." *Biochemical and Biophysical Research Communications* 529 (2): 257–62.
13. Fu, Laiyi, Lihua Zhang, Emmanuel Dollinger, Qinke Peng, Qing Nie, and Xiaohui Xie. 2020. "Predicting Transcription Factor Binding in Single Cells through Deep Learning." *Science Advances* 6 (51). https://doi.org/10.1126/sciadv.aba9031.
14. Gallo-Ramírez, Lilí Esmeralda, Alexander Nikolay, Yvonne Genzel, and Udo Reichl. 2015. "Bioreactor Concepts for Cell Culture-Based Viral Vaccine Production." *Expert Review of Vaccines* 14 (9): 1181–95.
15. Giri, L., H. Li, D. Sandgren, M. G. Feiss, R. Roller, B. C. Bonning, and D. W. Murhammer. 2010. "Removal of Transposon Target Sites from the Autographa Californica Multiple Nucleopolyhedrovirus fp25k Gene Delays, but Does Not Prevent, Accumulation of the Few Polyhedra Phenotype." *Journal of General Virology*. https://doi.org/10.1099/vir.0.024430-0.
16. Giri, Lopamudra, Michael G. Feiss, Bryony C. Bonning, and David W. Murhammer. 2012. "Production of Baculovirus Defective Interfering Particles during Serial Passage Is Delayed by

Removing Transposon Target Sites in fp25k." *The Journal of General Virology* 93 (Pt 2): 389–99.

17. Gotoh, Takeshi, Masamichi Fukuhara, and Ken-Ichi Kikuchi. 2008. "Mathematical Model for Change in Diameter Distribution of Baculovirus-Infected Sf-9 Insect Cells." *Biochemical Engineering Journal*. https://doi.org/10.1016/j.bej.2008.01.008.
18. Haas, Richard, and Lars K. Nielsen. 2005. "A Physiological Product-Release Model for Baculovirus Infected Insect Cells." *Biotechnology and Bioengineering* 91 (6): 768–72.
19. Hao, Wenrui, Elliott D. Crouser, and Avner Friedman. 2014. "Mathematical Model of Sarcoidosis." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1417789111.
20. Hernández Rodríguez, Tanja, Christoph Posch, Julia Schmutzhard, Josef Stettner, Claus Weihs, Ralf Pörtner, and Björn Frahm. 2019. "Predicting Industrial-Scale Cell Culture Seed Trains-A Bayesian Framework for Model Fitting and Parameter Estimation, Dealing with Uncertainty in Measurements and Model Parameters, Applied to a Nonlinear Kinetic Cell Culture Model, Using an MCMC Method." *Biotechnology and Bioengineering* 116 (11): 2944–59.
21. Hu, Yu-Chen, and William E. Bentley. 2000. "A Kinetic and Statistical-Thermodynamic Model for Baculovirus Infection and Virus-like Particle Assembly in Suspended Insect Cells." *Chemical Engineering Science*. https://doi.org/10.1016/s0009-2509(99)00579-5.
22. Jang, J. D., C. S. Sanderson, L. C. Chan, J. P. Barford, and S. Reid. 2000. "Structured Modeling of Recombinant Protein Production in Batch and Fed-Batch Culture of Baculovirus-Infected Insect Cells." *Cytotechnology* 34 (1-2): 71–82.
23. Jolliffe, I. T. 1986. "Principal Components Used with Other Multivariate Techniques." *Principal Component Analysis*. https://doi.org/10.1007/978-1-4757-1904-8_9.
24. Karr, Jonathan R., Jayodita C. Sanghvi, Derek N. Macklin, Miriam V. Gutschow, Jared M. Jacobs, Benjamin Bolival Jr, Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. 2012. "A Whole-Cell Computational Model Predicts Phenotype from Genotype." *Cell* 150 (2): 389–

401.

25. King, Andrew M. Q., Elliot Lefkowitz, Michael J. Adams, and Eric B. Carstens. 2011. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier.
26. Kirkwood, T. B., and C. R. Bangham. 1994. “Cycles, Chaos, and Evolution in Virus Cultures: A Model of Defective Interfering Particles.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.91.18.8685.
27. Kumar, Arun, and Michael L. Shuler. 1995. “Model of a Split-Flow Airlift Bioreactor for Attachment-Dependent, Baculovirus-Infected Insect Cells.” *Biotechnology Progress*. https://doi.org/10.1021/bp00034a008.
28. Kutzler, Michele A., and David B. Weiner. 2008. “DNA Vaccines: Ready for Prime Time?” *Nature Reviews. Genetics* 9 (10): 776–88.
29. Legastelois, Isabelle, Sophie Buffin, Isabelle Peubez, Charlotte Mignon, Régis Sodoyer, and Bettina Werle. 2017. “Non-Conventional Expression Systems for the Production of Vaccine Proteins and Immunotherapeutic Molecules.” *Human Vaccines & Immunotherapeutics* 13 (4): 947–61.
30. Licari, P., and J. E. Bailey. 1992. “Modeling the Population Dynamics of Baculovirus-Infected Insect Cells: Optimizing Infection Strategies for Enhanced Recombinant Protein Yields.” *Biotechnology and Bioengineering* 39 (4): 432–41.
31. Lindeberger, Christoph, Lukas Pflug, Holger Huebner, and Rainer Buchholz. 2012. “A Novel Model for Studying Baculovirus Infection Process.” *Biotechnology and Bioprocess Engineering*. https://doi.org/10.1007/s12257-011-0371-5.
32. Li, Shufen, Manli Wang, Shu Shen, Zhihong Hu, Hualin Wang, and Fei Deng. 2015. “The FP25K Acts as a Negative Factor for the Infectivity of AcMNPV Budded Virus.” *PloS One* 10 (5): e0128471.
33. Mirjalili, Seyedali. 2018. *Evolutionary Algorithms and Neural Networks: Theory and Applications*. Springer.

34. Moscardi, F. 1999. “Assessment of the Application of Baculoviruses for Control of Lepidoptera.” *Annual Review of Entomology* 44: 257–89.
35. Moscardi, Flavio, Marlinda Lobo de Souza, Maria Elita Batista de Castro, Mauricio Lara Moscardi, and Boguslaw Szewczyk. 2011. “Baculovirus Pesticides: Present State and Future Perspectives.” *Microbes and Microbial Technology*. https://doi.org/10.1007/978-1-4419-7931-5_16.
36. Oers, Monique M. van, Gorben P. Pijlman, and Just M. Vlak. 2015. “Thirty Years of Baculovirus-Insect Cell Protein Expression: From Dark Horse to Mainstream Technology.” *The Journal of General Virology* 96 (Pt 1): 6–23.
37. Orihara, Yutaka, Hiroshi Hamamoto, Hiroshi Kasuga, Toru Shimada, Yasushi Kawaguchi, and Kazuhisa Sekimizu. 2008. “A Silkworm Baculovirus Model for Assessing the Therapeutic Effects of Antiviral Compounds: Characterization and Application to the Isolation of Antivirals from Traditional Medicines.” *The Journal of General Virology* 89 (Pt 1): 188–94.
38. Power, J. F., and L. K. Nielsen. 1996. “Modelling Baculovirus Infection of Insect Cells in Culture.” *Cytotechnology* 20 (1-3): 209–19.
39. Power, J. F., S. Reid, K. M. Radford, P. F. Greenfield, and L. K. Nielsen. 1994. “Modeling and Optimization of the Baculovirus Expression Vector System in Batch Suspension Culture.” *Biotechnology and Bioengineering* 44 (6): 710–19.
40. Power, J., P. F. Greenfield, L. Nielsen, and S. Reid. 1992. “Modelling the Growth and Protein Production by Insect Cells Following Infection by a Recombinant Baculovirus in Suspension Culture.” *Cytotechnology* 9 (1-3): 149–55.
41. Puente-Massaguer, Eduard, Martí Lecina, and Francesc Gòdia. 2020. “Application of Advanced Quantification Techniques in Nanoparticle-Based Vaccine Development with the Sf9 Cell Baculovirus Expression System.” *Vaccine* 38 (7): 1849–59.
42. Qiu, Xiangguo, Jonathan Audet, Gary Wong, Stephane Pillet, Alexander Bello, Teresa Cabral, Jim E. Strong, et al. 2012. “Successful Treatment of Ebola Virus-Infected Cynomolgus Macaques

with Monoclonal Antibodies." *Science Translational Medicine* 4 (138): 138ra81.

43. Rappuoli, Rino, Mariagrazia Pizza, Giuseppe Del Giudice, and Ennio De Gregorio. 2014. "Vaccines, New Opportunities for a New Society." *Proceedings of the National Academy of Sciences of the United States of America* 111 (34): 12288–93.
44. Reddy, B., and J. Yin. 1999. "Quantitative Intracellular Kinetics of HIV Type 1." *AIDS Research and Human Retroviruses* 15 (3): 273–83.
45. Resmawan, Resmawan Resmawan, and Lailany Yahya. 2020. "Sensitifity Analysis of Mathematical Model of Coronavirus Disease (COVID-19) Transmission." *CAUCHY*. https://doi.org/10.18860/ca.v6i2.9165.
46. Roldão, António, Helena L. A. Vieira, Annie Charpilienne, Didier Poncet, Polly Roy, Manuel J. T. Carrondo, Paula M. Alves, and R. Oliveira. 2007. "Modeling Rotavirus-like Particles Production in a Baculovirus Expression Vector System: Infection Kinetics, Baculovirus DNA Replication, mRNA Synthesis and Protein Production." *Journal of Biotechnology* 128 (4): 875–94.
47. Saxena, Abha, Prasanna Kumar Byram, Suraj Kumar Singh, Jayanta Chakraborty, David Murhammer, and Lopamudra Giri. 2018. "A Structured Review of Baculovirus Infection Process: Integration of Mathematical Models and Biomolecular Information on Cell-Virus Interaction." *The Journal of General Virology* 99 (9): 1151–71.
48. Saxena, Abha, Suryateja Ravutla, Vikas Upadhyay, Soumya Jana, David Murhammer, and Lopamudra Giri. 2020. "Statistical Modeling of Cell-to-Cell Variability in Viral Infection during Passaging in Suspension Cell Culture: Application in Monte-Carlo Simulation." *Biotechnology and Bioengineering* 117 (5): 1483–1501.
49. Saxena, Abha, Vikas Upadhyay, Vaibhav Dhyani, Soumya Jana, and Lopamudra Giri. 2019. "Cell-to-Cell Variability in Protein Expression during Viral Infection: Monte-Carlo Simulation and Validation Based on Confocal Imaging." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in*

Medicine and Biology Society. Conference 2019 (July): 138–41.

50. Schiffer, Joshua T., David A. Swan, Amalia Magaret, Lawrence Corey, Anna Wald, Joachim Ossig, Helga Ruebsamen-Schaeff, et al. 2016. "Mathematical Modeling of Herpes Simplex Virus-2 Suppression with Pritelivir Predicts Trial Outcomes." *Science Translational Medicine* 8 (324): 324ra15.
51. Shan, Chao, Antonio E. Muruato, Bruno T. D. Nunes, Huanle Luo, Xuping Xie, Daniele B. A. Medeiros, Maki Wakamiya, et al. 2017. "A Live-Attenuated Zika Virus Vaccine Candidate Induces Sterilizing Immunity in Mouse Models." *Nature Medicine* 23 (6): 763–67.
52. Sidorenko, Y., and U. Reichl. 2004. "Structured Model of Influenza Virus Replication in MDCK Cells." *Biotechnology and Bioengineering* 88 (1): 1–14.
53. Singh, Ranjana, Surbhi Sharma, Venkatesh V. Kareenhalli, Lopamudra Giri, and Kishalay Mitra. 2020. "Experimental Investigation into Indole Production Using Passaging of E. Coli and B. Subtilis along with Unstructured Modeling and Parameter Estimation Using Dynamic Optimization: An Integrated Framework." *Biochemical Engineering Journal*. https://doi.org/10.1016/j.bej.2020.107743.
54. Society for Industrial and Applied Mathematics. 2002. *Proceedings of the Second SIAM International Conference on Data Mining*. Society for Industrial & Applied.
55. Srivastava, R., L. You, J. Summers, and J. Yin. 2002. "Stochastic vs. Deterministic Modeling of Intracellular Viral Kinetics." *Journal of Theoretical Biology* 218 (3): 309–21.
56. Volz, A., and G. Sutter. 2017. "Modified Vaccinia Virus Ankara: History, Value in Basic Research, and Current Perspectives for Vaccine Development." *Advances in Virus Research* 97: 187–243.
57. Wang, Junzhi. 2021. "New Strategy for COVID-19 Vaccination: Targeting the Receptor-Binding Domain of the SARS-CoV-2 Spike Protein." *Cellular & Molecular Immunology*.
58. Xue, Chuan, Avner Friedman, and Chandan K. Sen. 2009. "A Mathematical Model of Ischemic Cutaneous Wounds." *Proceedings of the National Academy of Sciences of the United States of*

America 106 (39): 16782–87.

59. Yao, Jason, Anna Pilko, and Roy Wollman. 2016. "Distinct Cellular States Determine Calcium Signaling Response." *Molecular Systems Biology* 12 (12): 894.
60. Zeb, Anwar, Ebraheem Alzahrani, Vedat Suat Erturk, and Gul Zaman. 2020. "Mathematical Model for Coronavirus Disease 2019 (COVID-19) Containing Isolation Class." *BioMed Research International* 2020 (June): 3452402.
61. Zhang, You-Hong, and José C. Merchuk. 2004. "A Mathematical Model of Baculovirus Infection on Insect Cells at Low Multiplicity of Infection." *Acta Biochimica et Biophysica Sinica* 36 (11): 729–40.
62. Zitzmann, Jan, Gundula Sprick, Tobias Weidner, Christine Schreiber, and Peter Czermak. 2017. "Process Optimization for Recombinant Protein Expression in Insect Cells." *New Insights into Cell Culture Technology*. https://doi.org/10.5772/67849.
63. Zwart, Mark P., Gorben P. Pijlman, Josep Sardanyés, Jorge Duarte, Cristina Januário, and Santiago F. Elena. 2013. "Complex Dynamics of Defective Interfering Baculoviruses during Serial Passage in Insect Cells." *Journal of Biological Physics* 39 (2): 327–42.

Author contributions: AS, KM, and LG were involved in conception and design. AS, LG, and acquired data. AS was involved in formulation of model. AS, and RS were involved writing the simulation code. AS, KM, DM, JC, and LG were involved in drafting and revising the article.

Competing interests: Authors declare that they have no competing interests.

Figure Legends:

Figure 1. Schematic for the network for virus internalization, replication, assembly and transport processes across the nucleus, cytoplasm, and extracellular region. (A) Representative image from transmission electron microscopy of virus-infected cells in training (blue) and testing (green) dataset. (B) Schematic diagram of replication cycle of baculovirus showing model variables. (C) The diagram shows

the network structure representation of transcription, translation, protein transport as well as virus packaging and assembly. The colored boxes show virus internalization (blue), DNA replication and RNA transcription (green), early protein synthesis (yellow), late protein synthesis (red), very late protein synthesis (violet), nucleocapsids and budded virus formation (magenta), and the occlusion (cyan) process. The description of model variables and rate constants are given in Tables 1 and 2, respectively.

Figure 2. Nucleocapsid formation and viral packaging for FP25K mediated transport of late protein and transcription of early protein Gp64. Schematic representation of the feedback structure for the four plausible models based on (A) hypothesis 1 (no feedback regulation), (B) hypothesis 2 (negative feedback regulation), (C) hypothesis 3 (positive feedback regulation), and (D) hypothesis 4 (negative and positive feedback regulation). Time course of model variables obtained from the simulation in single cells for (E) nucleocapsids in the nucleus, (F) nucleocapsids meant to become ODVs, (G) total ODVs in a cell, (H) unpackaged ODVs in the nucleus, (I) ODVs packaged inside polyhedral and (J) polyhedra in a cell. Validation of feedback regulation on model variables by Fp25k protein in single cells using two different scenarios (K) high level of Fp25k protein and (L) low level of Fp25k protein.

Figure 3. Clustering of infected cell population and corresponding variation in the rate of DNA replication. (A) Visualization of the infected cell population using principal component analysis. K-medoid clustering shows the presence of nine clusters in infected cells from various passages. (B) Box plot representation of the rate of DNA replication (kr_{DNA}) within each cluster.

Figure 4. Schematic diagram for the strategy used for parameter estimation in capturing cell to cell variability. (A) The infected cell characteristics were grouped in 9 clusters. The variability across clusters was attributed to (B) the rate of RNA synthesis $\left(ktr_{RNA}\right)$ which was chosen as a free parameter. The variability within a cluster was attributed to (C) the rate of viral DNA replication $\left(kr_{DNA}\right)$ which was chosen as another free parameter. (D) The remaining parameters of the model were kept as a fixed parameter. For kinetic parameter description, refer to Table 2.

Figure 5. Comparison of multiple model predictions over measured variables for training and testing data. The training dataset contains TEM image of virus-infected cells having distinct characteristics (A) cell plasma membrane (red arrow), nuclear membrane (green arrow) and virogenic stroma (yellow arrow), (B) polyhedra inside the nucleus (POL), (C) packaged nucleocapsids meant to become ODVs (VN_{VO}) (red), ODVs packaged inside polyhedra (VO_{POL}) (green), unpackaged ODVs in the nucleus ($VO_{,N}$) (yellow) and (D) unpackaged nucleocapsids in the nucleus ($VN_{,N}$). The violin plot of all the model variables from the fit results to single cell data and measured variables from TEM experiments for (E) number of unpackaged nucleocapsids in the nucleus ($VN_{,N}$) at 72hpi, (F) number of packaged nucleocapsids in ODV (VN_{VO}) at 72 hpi, (G) number of unpackaged ODVs in the nucleus ($VO_{,N}$), (H) number of packaged ODVs in polyhedra (VO_{POL}), (I) the total number of ODVs in a cell (VO) and (J) number of polyhedra in a cell (POL). The goodness of fit analysis for models M1, M2, M3, M4 in case of (K) $VN_{,N}$, (L) VN_{VO}, (M) $VO_{,N}$, (N) VO_{POL}, (O) VO and (P) POL. The TEM microscopy testing dataset contains virus-infected cells having distinct characteristics (Q) nucleocapsids meant to become BV in the cytoplasm (VN_{VB}) and (R) nucleocapsids at the budding site ($VN_{,M}$). Comparison of model 3 and 4 predictions over measured variables through violin plot representation of (S) number of nucleocapsids meant to become BV (VN_{VB}) and (T) number of nucleocapsids at the plasma membrane ($VN_{,M}$). E: Experiment, M1: Model 1, M2: Model 2, M3: Model 3, M4: Model 4. *p<0.05, **p<0.005, ***p<0.0005, NS: Non-significant.

Figure 6. Heat map representation of RMSE between experiment versus simulation results for each cell and model evolution. The x-axis represents the RMSE for each training (V1-V6) and testing (V7, V8) data of every cell on the y-axis for models (A) M1, (B) M2, (C) M3 and (D) M4.

Figure 7. Parameter sensitivity analysis of the model and categorization of parameters based on the level of sensitivity. The bars show the percentage changes of model variables (V1-V6) when varying

parameters by 20% (A) decrease and (B) increase in parameter from the basal value. For model variables and kinetic parameter description, refer to Table 1 and 2, respectively.

Figure 8. Response surface and corresponding contour plots based on the simulated data generated using factorial design. (A) Budded virus, (B) polyhedra production, (C) Gp64 protein expression and (D) polyhedrin protein expression as a function of MOI and TOH. Contours were created by depicting the response as a function of MOI and TOH and visualization of optimal conditions for the specific viral strain.

Table 1: Definition of model state variables.

S.no	Variable Symbol	Definition and unit
Virus Internalization		
1	$VP_{,Ex}$	number of virus particles in the extracellular region(#/cell)
2	$VP_{,M}$	number of virus particles on the cell surface(#/cell)
3	$VI_{,En}$	number of virus particles in endosome(#/cell)
4	$VN_{,C}$	number of nucleocapsids in the cytoplasm(#/cell)
5	$DNAv_{,N}$	number of uncoated viral DNA in the nucleus(#/cell)
Replication and Transcription		
6	$RNAe$	number of mRNA encoding early proteins(#/cell)
7	$DNA_{,N}$	number of DNA copies in the nucleus(#/cell)
8	RNA_l	number of mRNA encoding late proteins(#/cell)

9	RNA_{vl}	number of mRNA encoding very late proteins(#/cell)
Early protein synthesis		
10	$PGP_{,C}$	Gp64 protein in the cytoplasm ($\mu g/cell$)
11	$PGP_{,M}$	Gp64 protein at Plasma membrane ($\mu g/cell$)
Late protein synthesis		
12	$PCP_{,C}$	Capsid protein in the cytoplasm ($\mu g/cell$)
13	$PCP_{,N}$	Capsid protein in nucleus ($\mu g/cell$)
14	$PVP_{,C}$	vp39 protein in the cytoplasm ($\mu g/cell$)
15	$PVP_{,N}$	vp39 protein in nucleus ($\mu g/cell$)
16	$PFP_{,C}$	Fp25k protein in the cytoplasm ($\mu g/cell$)
17	$POE_{,C}$	ODV-E66 protein in the cytoplasm ($\mu g/cell$)
18	$POE_{,N}$	ODV-E66 protein in nucleus ($\mu g/cell$)
Very late protein synthesis		
19	$PPH_{,C}$	Polyhedrin protein in the cytoplasm ($\mu g/cell$)
20	$PPH_{,N}$	Polyhedrin protein in nucleus ($\mu g/cell$)
Nucleocapsids and budded virus formation		
21	$VN_{,N}$	number of Nucleocapsids in the nucleus (#/cell)
22	VN_{VB}	number of Nucleocapsids meant to become BV (#/cell)
23	VN_{VO}	number of Nucleocapsids meant to become ODV (#/cell)
24	$VN_{,M}$	number of Nucleocapsids at PM (#/cell)
25	VB	number of budded virus (#/cell)
Occlusion		
26	$VO_{,N}$	number of Occlusion Derived Virus in the nucleus (#/cell)
27	VO_{POL}	number of Occlusion Derived Virus packaged inside polyhedra (#/cell)
28	POL	number of polyhedra (#/cell)

Table 2: Definitions of model kinetic parameters and estimated value.

S.no	Parameter Symbol	Estimated value	Definition and unit
Virus transport			
1	$kvt_{Ex,M}$	9.23×10^{-1}	rate constant for virus transport from extracellular to the cell surface (plasma membrane (PM)) (h^{-1})
2	$kvt_{M,En}$	9.95×10^{-1}	rate constant for virus internalization from the cell surface (plasma membrane (PM)) (h^{-1})
3	$kvt_{En,C}$	9.99×10^{-1}	rate constant for nucleocapsid transport from endosome to cytoplasm (h^{-1})

4	$kvt_{C,N}$	9.69×10^{-1}	rate constant for nucleocapsid transport from cytoplasm to nuclear membrane and release of viral DNA into nucleus (h^{-1})
5	$kvt_{N,VS}$	8.93×10^{-1}	rate constant for the transport of viral DNA to virogenic stroma (h^{-1})
6	$kvt_{N,M}$	9.99×10^{-1}	rate constant for VP39 adided transport of nucleocapsid meant to become budded virus from the nucleus to PM
Protein transport			
7	$kpt_{C,M}$	8.21×10^{-1}	rate constant for the transport of protein from the cytoplasm to PM (h^{-1})
8	$kpt_{C,N}$	9.71×10^{-1}	rate constant for the transport of protein from the cytoplasm to the nucleus (h^{-1})
Protein association			
9	$kpa_{CP,DNA}$	9.98×10^{-1}	rate constant for association of capsid protein to DNA and assembly to form nucleocapsid (h^{-1})
10	$kpa_{GP,VN}$	8.44×10^{-1}	rate constant for association of Gp64 protein to nucleocapsids and budding from PM (h^{-1})
11	$kpa_{PH,VO}$	6.64×10^{-1}	rate constant for protein association of polyhedrin protein with ODV in nucleus (h^{-1})
12	$kpa_{VP,VN}$	1.32×10^{-1}	rate constant for association of vp39 protein to nucleocapsids (h^{-1})
13	$kpa_{OE,VN}$	4.19×10^{-2}	rate constant for association of ODV-E66 protein to nucleocapsids (h^{-1})
Virus assembly			
14	kva_{VO}	3.69×10^{-2}	rate constant for encapsulation of nucleocapsid in ODV-associated protein to form ODVs (h^{-1})
15	kva_{POL}	1.14×10^{-1}	rate constant for crystallization of polyhedrin protein to form polyhedra (h^{-1})
Replication and transcription			
16	kr_{DNA}	0.2-0.99	rate constant for replication of DNA (h^{-1})
17	ktr_{RNA}	0.4-0.99	rate constant for the transcription of RNA (h^{-1})
Translation			
18	ktl_{GP}	2.35×10^{-1}	rate constant for translation of Gp64 protein(h^{-1})
19	ktl_{CP}	9.99×10^{-1}	rate constant for translation of capsid protein (h^{-1})
20	ktl_{VP}	2.94×10^{-2}	rate constant for translation of vp39 protein (h^{-1})
21	ktl_{FP}	1.81×10^{-2}	rate constant for translation of Fp25k protein in cytoplasm (h^{-1})
22	ktl_{OE}	2.52×10^{-1}	rate constant for translation of ODV-E66 protein (h^{-1})
23	ktl_{PH}	1.98×10^{-1}	rate constant for translation of polyhedrin protein (h^{-1})
Feedback regulation			
24	knf_{GP}	7.27×10^{-1}	rate constant for negative feedback regulation by Fp25k protein on Gp64 protein transcription (h^{-1})
25	kpf_{OE}	9.99×10^{-1}	rate constant for positive feedback regulation by Fp25k protein

			on the transport of ODV-E66 protein from the cytoplasm to nucleus (h^{-1})
26	kpf_{PH}	9.99×10^{-1}	rate constant for positive feedback regulation by Fp25k protein on polyhedrin protein transcription (h^{-1})
Degradation			
27	kdv_{En}	7.27×10^{-2}	rate constant for virus degradation in endosome (h^{-1})
28	kd_{RNA}	4.46×10^{-1}	rate constant for degradation of RNA(h^{-1})
29	kd_{DNA}	6.28×10^{-1}	rate constant for degradation of DNA in nucleus (h^{-1})
30	$kdgp_{,C}$	1.28×10^{-1}	rate constant for degradation of Gp64 protein in cytoplasm (h^{-1})
31	$kdp_{,M}$	3.73×10^{-1}	rate constant for degradation of protein at PM (h^{-1})
32	$kdp_{,C}$	9.98×10^{-1}	rate constant for degradation of protein in cytoplasm (h^{-1})
33	$kdp_{,N}$	9.99×10^{-1}	rate constant for degradation of protein in nucleus (h^{-1})