Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-19 (2018) 1–4

Mathematical modeling as a tool to improve influenza vaccine production processes

Stefanie Duvigneau [*] **Robert Dürr** [**,***] **Tanja Laske** [***]
Mandy Bachmann [***] **Melanie Dostert** [***] **Udo Reichl** [*,***]
Achim Kienle [*,***]

[*] *Otto-von-Guericke-University, Magdeburg, Germany*
[**] *KU Leuven, Leuven, Belgium*
[***] *Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany*

Abstract:
Cell culture-based production of influenza vaccines is emerging as a promising alternative to conventional production in embryonated chicken eggs. Development and establishment of high-yield producer cell lines represents a major challenge to manufacture sufficient amounts of low-cost vaccines. One possible option to optimize vaccine production is to manipulate the expression of host cell factors relevant for virus replication. Lentiviral transduction is a gene editing method that allows to modify the expression of single or multiple host cell genes. However, due to different copy numbers and integration sites of the gene constructs the expression level shows a large cell-to-cell variability within the cell population. In this study, we will investigate the impact of genetic modifications on virus yield with the help of a structured population balance model. Therein, cell-to-cell variability is represented in terms of distributed kinetic parameter sets obtained after bootstrapping for five cell lines overexpressing a single gene. Moreover, we evaluate four different strategies to predict distributed parameter sets for cell lines overexpressing multiple genes based on the parameter distributions of the underlying single gene modifications. Furthermore, we will apply the most suitable prediction strategy to find a combination of gene modifications that leads to the highest virus productivity.

Keywords: population balance modeling, influenza vaccine production, multiscale modeling, cell-to-cell variability, genetically engineered cell lines

1. INTRODUCTION

Influenza is a highly contagious disease caused by different strains of influenza virus constituting a permanent threat to public health. The best way to counteract influenza is vaccination. Influenza vaccines consist of inactivated, attenuated or parts of influenza virus particles. A majority of these is produced in embryonated chicken eggs, which is a well established system with certain disadvantages, such as limited flexibility. To overcome drawbacks of egg-based processes animal cell cultures can be used as a substrate for virus production.

In this study, we used A549 cells as model cell line to produce influenza A virus particles. Genetic modifications of the cell line via lentiviral transduction techniques represent a promising opportunity to overcome cell specific bottlenecks. In Hamamoto et al. (2013) it was shown that overexpression or knock down of genes coding suitable host cell factors (HCF) may result in an increased virus yield. Experimental studies further indicate that manipulating the expression of multiple genes can result in even higher virus yields and are thus favored over single gene overexpressions (see Laske et al. (2018) and the references therein). However, predicting in advance which gene modifications will lead to a high virus yield is still a challenge and a large number of candidates need to be tested in time consuming and expensive experiments. This situation could be at least partially overcome with a model-based approach that uses a small number of experiments with cell lines overexpressing a single gene (SGOs) in order to infer information on the most promising cell lines with multiple gene overexpressions (MGOs). We selected single gene modifications where the gene products have shown interactions with the viral life cycle (Karlas et al., 2010). Furthermore, it has to be taken into account that the level of gene expression affected upon lentiviral transduction is not identical for all cells in a population and thus cell-to-cell variability is commonly observed. Lentiviral transduction causes this variability for different reasons. First, variable copy numbers and integration sites of a gene construct lead to heterogeneity in target gene expression between cells. Second, the number of gene candidates entered into the cells in the case of multiple gene modifications further affects this cellular heterogeneity. In this study, we only consider the first reason.

In the following, we will analyze the impact of selected gene candidates with regard to the amount of virus yield using a multiscale mathematical model of viral replication in cell cultures (Dürr et al., 2017). Directly including the variabil-

10.1016/j.ifacol.2018.09.004

ity caused by gene modification is challenging and requires a detailed mathematical description of cellular transcription and translation as well as of the molecular interactions of viral components with every cellular factor or pathway. Therefore, we assume that the variability introduced by genetic modifications can be expressed by the distribution of kinetic parameters. Those parameter distributions are generated by bootstrap parameter estimates and are used to account for cellular heterogeneity in a population balance modeling approach. The resulting population balance equation (PBE) represents a high dimensional partial integro differential equation and is solved numerically with a recently developed efficient approximate moment method (Dürr et al., 2017). Simulation results for the SGOs are validated with the help of experimental data. Furthermore, we will present four different strategies to generate distributed parameter sets for MGOs using the SGO parameter distributions. Virus titer measurements from infection experiments of MGOs with known genetic modifications are used to validate our simulation results. The last step comprises a combinatoric study, in which we use our approach to predict the virus yield for all possible MGOs. The most promising candidates, i.e. the ones with the highest yields, are shown and discussed.

2. MATHEMATICAL MODEL

Population balance modeling (Ramkrishna, 2000) is an established framework to account for heterogeneity in multicellular systems and is used here to describe the observed cell-to-cell variability with respect to the intracellular viral components and the kinetic parameters affected by the genetic modifications. The resulting model describes the virus production process on multiple scales which accounts for intra- and extracellular events. The intracellular level accounts for the major steps of the viral life cycle with focus on RNA replication and regulation (Heldt et al., 2013). The dynamics of the infected cell number density distribution $i_c(t, \mathbf{x})$ can be described by the following multidimensional population balance equation

$$\frac{\partial i_c}{\partial t} + \nabla_{\mathbf{x}^*} \left(\mathbf{h}^* \; i_c\right) = -(k_T^{\mathrm{Apo}} + k_i^{\mathrm{Apo}}) i_c + r^{\mathrm{inf}} \; T \,. \qquad (1)$$

Here, the extended single cell dynamics $\mathbf{h}^* = [\mathbf{h}, \mathbf{0}]^T$ of the extended state vector $\mathbf{x}^* = [\mathbf{x}, \mathbf{k}]^T$ describes the temporal evolution of the number density of cells depending on the different intracellular states $\mathbf{x}$ and the kinetic parameters $\mathbf{k}$. The single state vector $\mathbf{x}$ contains relevant intracellular compounds, such as viral proteins. In addition to the intracellular viral compounds, the parameters in the vector $\mathbf{k}$ vary within the population of infected cells as well, but are time invariant. More information is found in Dürr et al. (2017). Furthermore, the number of infected cells increases with rate proportional to r^{inf} and decreases with rate proportional to k^{Apo}. The apoptotic rate is divided into an apoptotic rate caused by infection (k_i^{Apo}) and a natural apoptosis rate (k_T^{Apo}). The infected cell dynamics is coupled to the temporal evolution of apoptotic cells I_a, target cells T, apoptotic target cells T_a and virus particles V. These species represent the extracellular level and their dynamics are described with ordinary differential equations, which are provided in Heldt et al. (2013). Since the PBE represents a high dimensional partial differential equation ($dim(\mathbf{x}^*) = 33$) that is coupled to a set of ODEs, standard numerical solution approaches like discretization based methods require a high numerical effort. For that reason, a recently developed approximate moment method is applied to achieve an efficient solution of the system (Dürr and Kienle, 2014; Dürr et al., 2017).

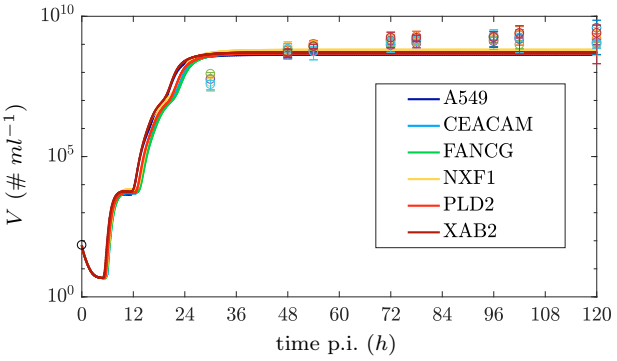

Fig. 1. *Experimental data and model simulations of virus yield for five different SGOs and the parental cell line (A549) at MOI* 10^{-4}.

3. SINGLE GENE OVEREXPRESSION CELL LINES

To determine the distributions for the parameters in $\mathbf{k}$ we used a single cell model of virus replication based on Heldt et al. (2013) and adapted it to experimental data of five SGOs as presented in Laske et al. (2018). We optimized the synthesis rates of viral mRNA (k_M^{Syn}), vRNA (k_V^{Syn}), cRNA (k_C^{Syn}), the binding rate of M1 to vRNP (k_{M1}^{Bind}) as well as the release (k^{Rel}) and import rate of the virus (k^{Imp}) using the fSSm algorithm (Egea et al., 2007). The optimization procedure was repeated 1000 times by randomly generating an *in silico* data set using the experimental error (Laske et al., 2018). The resulting bootstraped parameter distributions are used to simulate the population balance model with $T(0) = 7 \cdot 10^5 \# \; ml^{-}1$ initial target cells and a multiplicity of infection $MOI = V(0)\, T^{-1}(0) = 10^{-4}$. The resulting virus dynamics for each cell line is shown in Fig. 1. The small discrepancy between the shown low-MOI experimental data and simulation can be explained by the fact that the bootstrap parameter distributions have been estimated from high-MOI datasets.

4. DISTRIBUTION STRATEGIES FOR MGOS

In contrast to SGOs, in MGOs more than one gene is overexpressed. We assume that parameter distributions of MGOs can be determined by combining the corresponding parameter distributions of the underlying SGOs. However, it is not known how the SGO parameter distributions should be combined to make valid predictions. For that reason, we compare four strategies to construct MGO parameter distributions from the SGO parameter distributions and validate those with measured virus yields of four specific MGOs (see Laske et al. (2018) for details). In the following, we select the parameter distributions and virus dynamics of MGO 3 as exemplary MGO to illustrate the different outcomes of the strategies (Fig. 2 and Fig. 3).

Low impact strategy: For this strategy, we assume further gene overexpressions have only a low impact to the viral

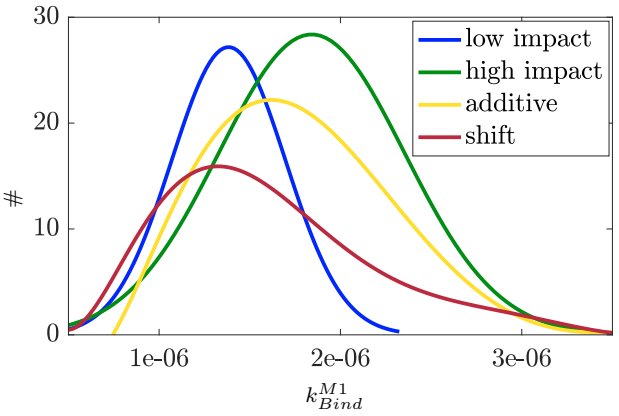

Fig. 2. *Distribution of the kinetic parameter* k^{M1}_{Bind} *of MGO 3* for all evaluated strategies

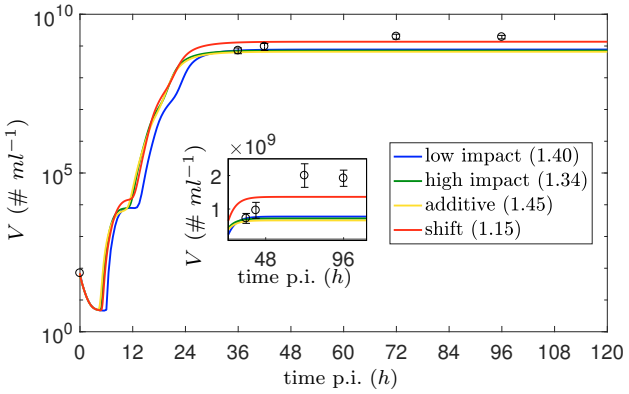

Fig. 3. *Experimental data for MGO 3 and corresponding model simulations for presented strategies (overall RMS values in brackets).*

replication cycle. For that reason, the parameter distributions for the MGO are described by a weighted sum of five logarithmic Gaussians located between the median values of the first SGO (in the following called base) and the median values of the SGO closest to the base value

$$k_{\mathrm{i}} \sim \sum_{l=1}^{5} a_l \ \mu_l \ e^{\mathcal{N}(0,\sigma_l)} \,. \tag{2}$$

The Gaussians are weighted with the parameter a_l. A former study demonstrated the impact of different shaped parameter distributions (Dürr et al., 2017). Due to the results in Dürr et al. (2017), the scaling factors a_l for all distributions are selected as $[a_1,\ a_2,\ a_3,\ a_4,\ a_5] = [0.05,\ 0.3,\ 0.3,\ 0.3,\ 0.05]$.

High impact strategy: In a desirable scenario, further gene modifications have a high impact on parameter distributions. Therefore, parameters of the MGOs are distributed between the median values of the base and the median values of the SGO which cause the greatest change in the parameter distribution. The distributions are created as a weighted sum of five logarithmic Gaussians similar to the low impact strategy.

Additive strategy: The usage of parameter distributions obtained from SGO experiments can lead to more precise parameter distributions for the MGOs. The overlay of the SGO parameter distributions is achieved by selecting the highest absolute frequency for a certain parameter value to receive an MGO parameter distribution. The order of the gene overexpressions which are performed after the first gene modification is unknown. After comparison of the low and high impact strategy, we found that the high impact approach showed best agreement to the experimental data. Therefore, the median value of each parameter with the greatest absolute distance to the base median value is the selection criterion for the overlayed parameter distribution. A grid with 150 bins between the bounds of the parameter distributions is defined to obtain the resulting overlay for each parameter. Each SGO distribution is multiplied by the relative overexpression level of the corresponding gene to weight the possible influence of each gene modification in an MGO.

Shift strategy: The last strategy accounts for both parameter distributions and the corresponding relative overexpression levels of all gene candidates. In contrast to the preceding strategies this strategy deals with the mean value and the standard deviation of each SGO parameter distribution to shift the histogram of the base and construct the MGO parameter distributions. Similar to the additive strategy, we chose the distribution of the gene modification with the highest distance to the median value of the first SGO to shift the parameter distribution for the MGO.

Strategy evaluation: From Fig. 3 we can obtain that all strategies are able to capture the trend of the experimental data. To compare the quality of the different strategies, the root mean square (RMS) between experimental and simulated data was calculated on a logarithmic scale for all MGOs (overall RMS in Fig 3) indicating that the shift strategy is the most suitable approach.

5. MGO COMBINATION STUDY

In the previous section, we showed that the shift strategy is a suitable method to reconstruct parameter distributions of MGOs from SGO parameter distributions. An application of this strategy is to investigate which possible combinations of the available SGOs are most promising by means of obtaining a high virus yield. This represents an alternative to an expensive "brute force" experimental study for such a high number of combinations. In contrast to the evaluation of the shift strategy in the previous section, a selection of one distribution based on the distances of the medians is not necessary since we focus on combinations of our given SGOs to construct new MGO parameter distributions. Hence, we have simulated the virus dynamics for two to five SGO combinations with the multiscale model. Finally, 320 MGO candidates have been studied numerically.

A detailed analysis of the stepwise inclusion of each gene modification revealed that more modifications in a proper sequence can lead to higher yields of virus particles. A comparison of the best results of each combination structure (overexpression of either two, three, four or five gene candidates) revealed that NXF1 is part of each MGO which showed an ten-fold increase in the virus concentration at 48 h compared to the best SGO. Fig. 4 shows the virus dynamics of the SGO overexpressing NXF1 in comparison to cell lines with two, three, four and five gene modifications and NXF1 as first gene modification.

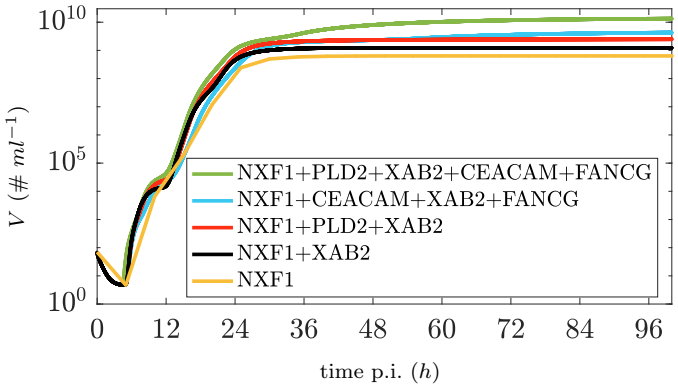

Fig. 4. *Simulated virus dynamics for the best combinations with NXF1 as base.*

6. CONCLUSION

In this paper, cell-to-cell variability caused by genetic modifications in cell lines for influenza vaccine production has been investigated. This variability can be incorporated by assuming a relationship between a heterogeneity caused by gene modification and parameter distributions of our model for the viral replication cycle. These distributions were taken into account by a population balance modeling approach. Therein, different distributed parameter sets have been used to represent cell lines with gene modifications. One goal of this study was the evaluation of different strategies to generate parameter distributions of producer cell lines overexpressing multiple host cell genes upon lentiviral transduction. For that purpose, we first compared four approaches to determine such distributions. With respect to the experimental reference data, the shift strategy seems to be best for most of the investigated MGOs. However, we need both, parameter bootstraps and the expression level of the genes.

Subsequently, the shift strategy was used to investigate all possible MGO combinations with regard to their virus yield. The most promising MGO candidate, which overexpresses five genes simultaneously, showed a 10-fold higher virus yield compared to the best SGO overexpressing NXF1. In addition, also other MGO candidates with less than five overexpressed genes produced higher virus titers compared to SGOs and all included NXF1 overexpression as first gene modification. Since inhibition of NXF1 leads to less mRNA coding for viral hemagglutinin and neuraminidase in the cytoplasm in A549 cells (Larsen et al., 2014), the overexpression could increase the amount of such mRNAs. Considering a robust transcription, the higher amount of mRNAs results in an increased quantity of the viral surface proteins. This might be a key mechanism for the increase in virus concentration, since the modification guarantees a sufficient supply of virus surface protein such that release of viral progeny is improved. This is in agreement with other model predictions, which determined the viral release as a kinetic bottleneck (Laske et al., 2018).

In further studies we want to validate our theoretical results with additional experimental results. For that purpose, more gene candidates have to beevaluated that show a higher impact on the virus yield. However, such experimental investigations are rather challenging as it is mostly not known beforehand if a modification of a cellular factor affects the virus life cycle. This missing biological knowledge is the major bottleneck in the iterative loop of experimental and model-based theoretical investigations and has to be overcome to enable a profitable use of the presented application. Yet, if promising SGOs are available, our approach provides a helpful tool for a facilitated selection of gene modifications in experimental planning. Thereby, the number of screening experiments can be reduced leading to reduced costs and the chance of finding high yield cell lines is increased.

ACKNOWLEDGEMENTS

This work was supported by the funding of the German Ministry of Education and Research (BMBF) as part of the e:Bio project CellSys (grant number 031 6189 A), the EU-programme ERDF (European Regional Development Fund) within the research center of dynamic systems (CDS) and the International Max Planck Research School for Advanced Methods in Process and Systems Engineering (IMPRS ProEng) in Magdeburg.

REFERENCES

Dürr, R. and Kienle, A. (2014). An efficient method for calculating the moments of multidimensional growth processes in population balance systems. *The Canadian Journal of Chemical Engineering*, 92, 2088–2097.

Dürr, R., Müller, T., Duvigneau, S., and Kienle, A. (2017). An efficient approximate moment method for multidimensional population balance models - Application to virus replication in multi-cellular systems. *Chemical Engineering Science*, 160, 321–334.

Egea, J., Rodríguez-Fernández, M., Banga, J., and Mart, R. (2007). Scatter search for chemical and bio-process optimization. *Journal of Global Optimization*, 37(3), 481503.

Hamamoto, I., Takaku, H., Tashiro, M., and Yamamoto, N. (2013). High yield production of influenza virus in Madin Darby Canine Kidney (MDCK) cells with stable knockdown of IRF7. *PloS One*, 8(3), e59892.

Heldt, F.S., Frensing, T., Pflugmacher, A., Gröpler, R., Peschel, B., and Reichl, U. (2013). Multiscale modeling of influenza A virus infection supports the development of direct-acting antivirals. *PLoS Computational Biology*, 9(11).

Karlas, A., Machuy, N., Shin, Y., Pleissner, K., Artarini, A., Heuer, D., Becker, D., Khali, H., L.A., O., Hess, S., Mürer, A., Müller, E., Wolff, T., Rudel, T., and Meyer, T. (2010). Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*, 463, 818–822.

Larsen, S., Bui, S., Perez, V., Mohammad, A., Medina-Ramirez, H., and Newcomb, L. (2014). Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export. *Virology Journal*, 11(154).

Laske, T., Bachmann, M., Dostert, M., Frensing, T., Wirth, D., Karlas, A., Jordan, I., Kienle, A., and Reichl, U. (2018). Model-based analysis of influenza A virus production in genetically engineered cell lines reveals changes in early dynamics of virus replication. *in preparation*.

Ramkrishna, D. (2000). *Population Balances: Theory and Applications to Particulate Systems in Engineering*. Academic Press, San Diego.