TU Delft

Delft University of Technology

Using artificial neural networks to accelerate flowsheet optimization for downstream process development

Keulen, Daphne; van der Hagen, Erik; Geldhof, Geoffroy; Le Bussy, Olivier; Pabst, Martin; Ottens, Marcel

DOI
10.1002/bit.28454

Publication date
2023

Document Version
Final published version

Published in
Biotechnology and Bioengineering

Citation (APA)
Keulen, D., van der Hagen, E., Geldhof, G., Le Bussy, O., Pabst, M., & Ottens, M. (2023). Using artificial neural networks to accelerate flowsheet optimization for downstream process development. *Biotechnology and Bioengineering*, *121 (2024)*(8), 2318-2331. https://doi.org/10.1002/bit.28454

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Check for updates

Received: 27 February 2023 | Revised: 11 May 2023 | Accepted: 17 May 2023

DOI: 10.1002/bit.28454

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Using artificial neural networks to accelerate flowsheet optimization for downstream process development

Daphne Keulen[1] | Erik van der Hagen[1] | Geoffroy Geldhof[2] | Olivier Le Bussy[2] | Martin Pabst[1] | Marcel Ottens[1]

[1]Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

[2]GSK, Technical Research & Development—Microbial Drug Substance, Rixensart, Belgium

Correspondence
Marcel Ottens, Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.
Email: m.ottens@tudelft.nl

Funding information
GlaxoSmithKline Biologicals S.A. (Belgium)

Abstract

An optimal purification process for biopharmaceutical products is important to meet strict safety regulations, and for economic benefits. To find the global optimum, it is desirable to screen the overall design space. Advanced model-based approaches enable to screen a broad range of the design-space, in contrast to traditional statistical or heuristic-based approaches. Though, chromatographic mechanistic modeling (MM), one of the advanced model-based approaches, can be speed-limiting for flowsheet optimization, which evaluates every purification possibility (e.g., type and order of purification techniques, and their operating conditions). Therefore, we propose to use artificial neural networks (ANNs) during global optimization to select the most optimal flowsheets. So, the number of flowsheets for final local optimization is reduced and consequently the overall optimization time. Employing ANNs during global optimization proved to reduce the number of flowsheets from 15 to only 3. From these three, one flowsheet was optimized locally and similar final results were found when using the global outcome of either the ANN or MM as starting condition. Moreover, the overall flowsheet optimization time was reduced by 50% when using ANNs during global optimization. This approach accelerates the early purification process design; moreover, it is generic, flexible, and regardless of sample material's type.

KEYWORDS
artificial neural networks, chromatography, downstream process development, flowsheet optimization, mechanistic modeling, model-based process optimization

1 | INTRODUCTION

Purifying biopharmaceuticals is crucial to reduce contaminants to very low levels, which ensures safety and efficacy of the product. The downstream process consists of a combination of multiple separation techniques such as filtration, centrifugation, and chromatography. Chromatography is a powerful separation technique and has been employed in the industrial bioprocesses for decades; it is generally the most essential technique to achieve high product purity (Łącki, 2018). A purification process is developed by employing a certain approach, for example, trial-and-error, design of experiments (DoE), or modeling based. An overview of these different

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Biotechnology and Bioengineering* published by Wiley Periodicals LLC.

downstream process development strategies and recent advancements has been described elsewhere (Keulen et al., 2022). DoE is based on statistical methods and most commonly applied for process development in pharmaceutical industry (Bhambure & Rathore, 2013; Stamatis et al., 2019). It provides a multidimensional model that correlates the effects of various factors on the critical quality attributes (CQA). CQA is an essential aspect of the Quality-by-Design (QbD) guidelines, which is a strategy of process development to ensure quality and performance of the final product (FDA, 2004; ICH, 2009; Yu, 2008). As statistical methods provide little process-understanding and extrapolation is not possible, DoE is inadequate for overall process optimization. Therefore, the pharmaceutical industry is shifting toward a model-based process development strategy that is compliant with the QbD guidelines and with the adoption of Industry 4.0. Industry 4.0 desires a full digitalization of the whole manufacturing process; monitored data are collected and communication between machines could directly improve the process (Chen et al., 2020; Portela et al., 2020; Reinhardt et al., 2020; Silva et al., 2020). In this new era, model-based techniques are essential, involving mathematical mechanistic models (MMs), hybrid modeling, and artificial intelligence (AI). MMs are based on physical correlations and attempt to describe the real process (Felinger & Guiochon, 2004). The combination of AI techniques with mechanistic modeling could eliminate shortcomings in both techniques, and so improve the applicability and usability (Lin et al., 2021; Narayanan et al., 2021; von Stosch et al., 2014). The potential of applying AI-driven models for process development and their practical implementations have been discussed elsewhere (Rathore et al., 2023; von Stosch et al., 2021).

Developing a purification process requires making decisions such as type and sequential order of purification techniques, operating conditions, and costs (Huuk et al., 2014; Nfor et al., 2013). Minor variations in operating conditions may critically impact the performance of subsequent purification steps. In addition, it should be noted that the most optimal purification process may not consist of each unit operation performing at its individual optimum. Hence, to find the optimal purification process, it is pivotal to optimize the entire purification sequence at once by screening the overall design space. The optimal purification process is defined by certain process performances such as, yield, purity, productivity, or buffer consumption. However, for early process designs, the type and order of unit operations have yet to be decided. Superstructures contain all possible process configurations, each process configuration is also referred as flowsheets. Flowsheet optimization evaluates each flowsheet to find the optimal sequence for purifying the product, which can support decision-making on early process designs (Yeomans & Grossmann, 1999).

Kawajiri (2021) described different optimization strategies for chromatographic modeling and summarized related studies. Moreover, an open-software optimization framework for modeling conventional and advanced batches and continuous chromatography processes was developed by Schmölder and Kaspereit (2020). However, both studies are applicable to batch or continuous chromatography, but not to flowsheet optimization. Nfor et al. (2013) applied a top-to-bottom optimization approach to obtain a minimum number of purification steps in the final process. As sequential optimization can lead to a suboptimal process, Huuk et al. (2014) simultaneously optimized a two-step ion-exchange chromatography process. A similar approach was applied by Pirrung et al. (2019) simultaneously optimizing an integrated process of three chromatographic steps (e.g., cation exchange, hydrophobic interaction, and mixed-mode) including buffer exchange steps if needed (e.g., ultra- and diafiltration).

Many parameters play a role at an early-stage-design, for instance the number, order, and type of unit operations and their operating conditions. Finding global optima is therefore a complex task. The main aim of flowsheet optimization, for early process design, is to find the most effective sequence unit operation(s) and an estimation of their operating conditions. MMs are very appropriate for flowsheet optimization because of their extrapolation capabilities. However, these models can be speed-limiting when used for optimization purposes and therefore, using meta-models, such as artificial neural networks (ANN), as a representation of the MM can accelerate the optimization. In the early 2000s, Nagrath et al. (2004) already established a hybrid model optimization framework for preparative chromatography, using ANNs for speed improvement. In the work of Pirrung et al. (2017, 2019), all flowsheets of a superstructure were evaluated by a global and local optimizer; the outcomes of the global optimizer was used as starting conditions for the local optimizer. In this case, ANNs replaced the mechanistic model during global optimization; however, these ANNs were less precise and therefore unable to always find realistic results. The local optimization took around 80% of the total optimization time. Another approach would be to focus on the global optimization and to first find the most promising sequence(s) of unit operations, and only optimize a selection of best processes locally. In this way the number of flowsheets to be evaluated during local optimization would be significantly reduced and so the overall optimization time. To realize this, ANNs that function as surrogate models should be developed and therefore additional input parameters, the mass of each protein, are needed. However, increasing the number of input parameters for the ANN makes it more challenging to generate accurate ANNs with a limited number of sample points.

In this approach, we performed a flowsheet optimization to evaluate the use of ANNs versus MMs in identifying the overall best process sequence(s) during global optimization. The most promising process options can be optimized locally, hence saving a time-consuming task in which no significant better process is obtained. First, we developed ANNs for each chromatography mode and evaluated their accuracy in terms of R^2 and root mean squared error (RMSE) values. Secondly, we created a superstructure optimization framework in which MM and/or ANNs were used. At last, we evaluated, in terms of time and precision, if and when ANNs would be sufficient for flowsheet optimization purposes. We compared two optimization frameworks in which only a selection of best processes was evaluated locally: (i) global and local optimization using MMs and (ii) global optimization using ANNs and local optimization using MMs.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | MATERIALS AND METHODS

2.1 | Flowsheet optimization workflow

In this study a superstructure of three different chromatography modes in a maximum sequence of three unit operations was evaluated. Only flowsheets satisfying certain conditions are considered; for example, at least one unit operation is needed for the purification. To generate a maximum number of structures that confirm defined conditions, this problem is mathematically formulated as

$$y = [y_1, y_2, \ldots, y_n], \tag{1}$$

$$\text{s. t.} \quad \sum y \geq 1, \tag{2}$$

$$y_i \neq y_j \ \text{ for all } \ y_i > 0. \tag{3}$$

$$\text{For } i = 2, 3, \ldots, n:$$

$$\text{if } \ y_i > 0, \ \text{ then } \ y_{i-1} > 0, \tag{4}$$

where y is the process configuration, in which n, in this case $n = 3$, is the length of the vector. The variable $y_i \in \{0, 1, 2, 3\}$ represents the value of the ith element of vector y. The first statement, Equation (1), defines the set of all possible vectors y, where each element is number between 0 and 3, which in this study represents the considered chromatography modes, none, CEX, AEX, and HIC, respectively. The second statement assures, Equation (2), to have at least one unit operation present in the sequence. The third statement, Equation (3), ensures that each mode can only appear once in the sequence. The conditional constraint in Equation (4) is applied to all positions in the sequence, except the first position. This constraint imposes that any occupied position in the sequence must be preceded by another occupied position. This ensures to have no isolated modes in the sequence, and requiring all modes to be linked. For example, $y = [1, 3, 0]$ is a two-step chromatography process of CEX followed by HIC. This mathematical formulation can be easily extended for more and different types of unit operations.

Each flowsheet of the superstructure was optimized according to certain objective(s) and constraint(s), these are described in Section 2.5. The objective is to find an initial concept of a purification process. Therefore, we focused on the global optimization to select the best processes. A minor local optimization was performed afterwards, as at the early stage of the local optimization, using the Nelder–Mead algorithm, the steps can be larger towards the local minimum and therefore the solution is already closer to the final minimum. Subsequently, the selected process(es) were further optimized locally using MMs. The outcome of the foregoing global and local optimization was used as initial guess for the final local optimization. The overall flowsheet optimization workflow is shown in Figure 1, in which framework A runs the global optimization and minor local optimization using the MM, while framework B uses the ANN. In this way a fair comparison can be made between using MMs or ANNs during the flowsheet optimization.

After the global and minor local optimization according to the set objective, we used the weighted overall performances (WOP) to select the "best processes." The WOP was determined as

$$\begin{aligned} \text{WOP} = {} & 0.5 \times \text{purity} + 0.3 \times \text{yield} \\ & + 0.2 \times (100 - \text{buffer consumption}), \end{aligned} \tag{5}$$

where the purity (%) is determined by dividing the amount of product by the total amount of proteins present in the product pool. The yield (%) is determined by the total amount of product recovered divided by

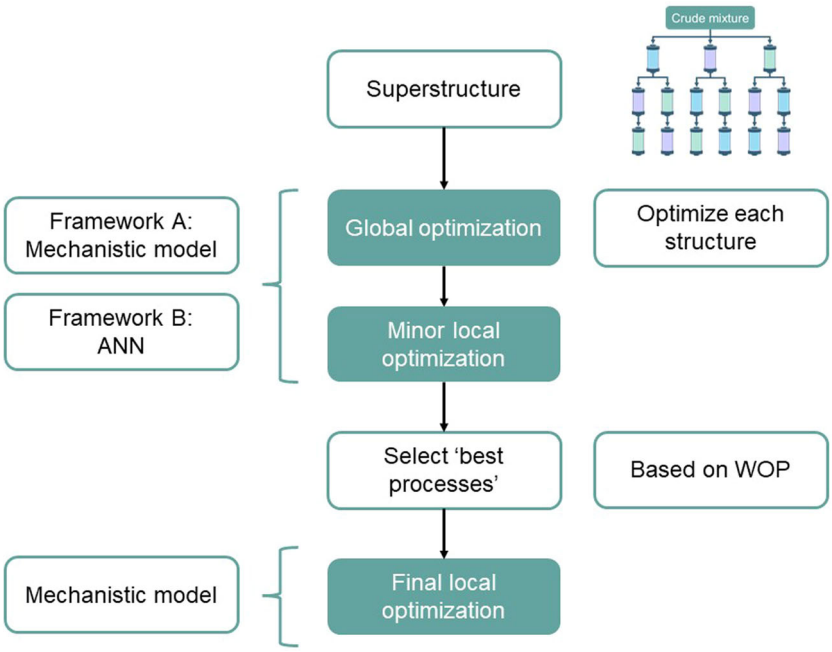


FIGURE 1 Each flowsheet of the superstructure, upper right figure, is first optimized globally to select the "best processes." These are further optimized using a final local optimizer. Framework A used mechanistic models (MMs) and framework B used artificial neural networks (ANNs) for global optimization. WOP, weighted overall performances.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the loaded amount of product. The buffer consumption ($L/g_{product}$) is approximately between 1 and 50. Subtracting the buffer consumption from 100 ensures the WOP increases when less buffer is consumed. Here, 100 is chosen to be in a similar range as the purity and yield.

Two other requirements in both global and local optimizers were:

- The next-unit operation could only be evaluated if the previous unit operation achieved a yield higher than 5%. This prevents the solver from failing because of too low concentration values.
- Between two unit operations, it was required to adapt the salt concentration to the conditions of the subsequent unit operation.

2.2 | Chromatography mechanistic model

To describe the dynamic adsorption behavior in the chromatographic process, we used the equilibrium transport dispersive model in combination with the linear driving force as follows:

$$\frac{\partial C_i}{\partial t} + F\frac{\partial q_i}{\partial t} = -u\frac{\partial C_i}{\partial x} + D_{L,i}\frac{\partial^2 C_i}{\partial x^2}, \tag{6}$$

$$\frac{\partial q_i}{\partial t} = k_{ov,i}\left(C_i - C^*_{eq,i}\right), \tag{7}$$

$$k_{ov,i} = \left[\frac{d_p}{6k_{f,i}} + \frac{d_p^2}{60\varepsilon_p D_{p,i}}\right]^{-1}, \tag{8}$$

where C and q are the concentrations in the liquid and solid phase respectively, and $C^*_{eq,i}$ is the liquid phase concentration in equilibrium with the solid phase. F is the phase ratio, defined as $F = (1 - \varepsilon_b)/\varepsilon_b$, where ε_b is the bed porosity. The interstitial velocity of the mobile phase is represented by u, and the axial dispersion coefficient by D_L. t and x indicate the time and space, respectively. $k_{ov,i}$ is the overall mass transfer coefficient defined as a summation of the separate film mass transfer resistance and the mass transfer resistance within the pores (Ruthven, 1984). Here, d_p is the particle diameter, ε_p is the intraparticle porosity, and D_p is the effective pore diffusivity coefficient. The first term represents the film mass transfer resistance, $k_f = D_f Sh/d_p$, in which D_f is the free diffusivity and Sh is the Sherwood number. More information on the MM can be found in a previous study (Nfor et al., 2011). Moreover, we used the multicomponent mixed-mode isotherm, as formulated by Nfor et al. (2010) and described in Appendix B.

2.3 | Developing artificial neural networks

A complete ANN consists of multiple layers of interconnected nodes, also known as artificial neurons, in which each neuron of one layer is connected with each neuron in the next layer (Madden et al., 2001). The outcome of each neuron is calculated by its activation function (σ), which is determined by function (z). Commonly used activation functions are Rectified Linear Unit (ReLU), sigmoid, and tangens hyperbolicus (Müller & Guido, 2017). The function (z) is determined by the weighted sum (w) of their inputs (x) added with a bias (b). The overall outcome of a neuron is mathematically represented as

$$\sigma(z) = \sigma\left(\sum_{i=1}^{j} w_i \cdot X_i + b\right), \tag{9}$$

where j is the number of neurons for the previous layer and σ represents the activation function. The neural network is trained by minimizing the error between the predicted and target output, this can be achieved by adjusting the weight and bias parameters of each neuron. In this work we used a deep neural network consisting of an input layer, two or three hidden layers, and an output layer. Determining the number of hidden layers, and other hyperparameters (e.g., batch size, number of epochs, and neurons), was done by varying the hyperparameter values and evaluating the effect on the ANN's accuracy. We used a ReLU activation function for the hidden layers as it is computationally more efficient; the sigmoid activation function was used for the output layer (Nwankpa et al., 2018).

The chromatographic MM performed numerous simulations to generate data that can be used for creating the neural network. The chosen input variables are the mass of each component, amount of loading, gradient length, initial and final salt concentrations, and the lower and upper cut points in percentage of the peak maximum (Figure 2). To model a sequence of unit operations, the mass of each protein, volume, and salt concentration present in the product pool are needed as input for the next unit operation. The mass in the product-pool varies and thus the mass as input for the next unit operation also varies. Therefore, the mass of each protein is needed as an input parameter for the ANNs. All output variables were taken from the product pool; mass of each component, volume, salt concentration and each cut point in column volume (CV). We noticed that including salt concentration of the product pool as an output variable increased the ANN's accuracy. We used the salt concentration as an output variable, but this value can also be calculated using the initial and final salt concentration (input parameters) and the cut points in CV (output parameter). Including the cut points in CV as output, resulted also in a better prediction of the volume.

We used the Latin hypercube sampling method of the pyDOE package for generating randomized data. The parameter space was based on prior-knowledge of biopharmaceutical downstream processes (Fellner et al., 2003). This was applied to both input and output parameters and minimized the "black-box" size. The best ANN was chosen out of 10 trained ANNs, as each time the weight and biases are trained in a different way and therefore the accuracy can differ. Moreover, the data were divided into 70% training, 15% validation, and 15% testing data. All other settings are described in Section 2.4. The used hyperparameters for each ANN of each chromatography mode are given in Table 1.

The ANN performance was assessed by the R^2 and RMSE value; these are based on the values predicted by the MM and calculated as follows:

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

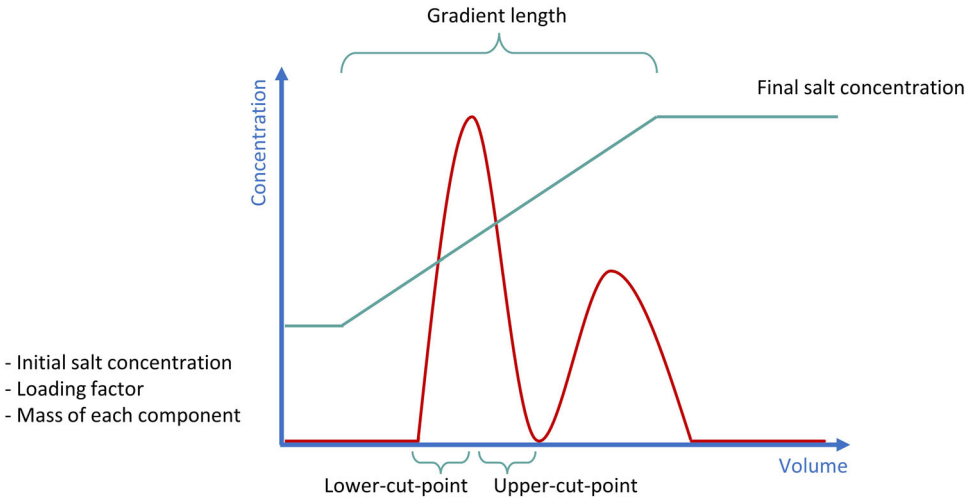


FIGURE 2 Input parameters used for the artificial neural network (ANN). Initial salt concentration ranging from 5 to 300 mM for CEX and AEX, and 100 to 500 mM for HIC. The final salt concentration ranging from 100 to 1200 mM for CEX and AEX, and 5 to 300 mM for HIC. The gradient length in the range of 1–10 column volume (CV), and loading factor from 0.05 to 5 CV. The concentration, converted to mass, ranging from 0.001 to 4 g/L. Both cut points in percentage of the peak maximum, lower cut point from 1% to 80% and upper cut points from 20% to 99%.

TABLE 1 Overview of final hyperparameters for each chromatography mode.

Hyperparameter	CEX	AEX	HIC
Batch size	512	128	512
Epochs	100	200	100
Number of hidden layers	2	3	2
Number of neurons	50	50	50
Learning rate	0.01	0.01	0.01

$$R^2 = 1 - \frac{\sum_i^N (y_i - Y_i)^2}{\sum_i^N (Y_i - \bar{Y}_i)^2}, \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{\sum_i^N (y_i - Y_i)^2}{N}}, \tag{11}$$

where Y_i is the mechanistic model data and y_i the data predicted by the ANN. N is the amount of data points used, and the $\overline{Y_i}$ is the mean of all the mechanistic model data points. Moreover, plots of the residual values are provided to show the data's randomness. The R^2 is a relative measure of fit and represents the proportion of variance explained by the relation between two variables. While the RMSE value is an absolute measure of fit that indicates the absolute mean difference between the predicted and true values.

2.4 | Numerical methods

All codes are written in Python (version 3.7), which is a free and open-source programming language. An overview of the used python libraries is given in Appendix A.

2.4.1 | Dynamic chromatography column model

The Method of Lines is applied for the spatial discretization to transfer partial differential equations (PDEs) into ordinary differential equations (ODEs) with respect to time. Moreover, a fourth-order central difference scheme for both first- and second-order derivatives with respect to space are used. The system of ODEs is solved using the LSODA (Livermore Solver for Ordinary Differential Equations) algorithm from the scipy.integrate package. This method automatically switches between the nonstiff Adams method and the stiff BDF method (Petzold, 1983).

2.4.2 | Optimization

The scipy.optimize package was used for the optimization; the differential_evolution algorithm for the global optimization and Nelder–Mead algorithm for the local optimization. The maximum number of iterations for global optimization was 9 and the population size 10 when using MMs, and for ANNs maxiter was 15 and population size was 20. Latin hypercube sampling was used to initialize the population. The maximum number of iterations for local optimization was 20. The relative tolerance for global and local optimization was 1e−2, and the function tolerance 1e−2. The maximum number of iterations for the final local optimization was 200. Due to limited accuracy of the ANNs, the mass could be predicted higher or lower, and so influencing the performance measurements. The predicted mass was set to the mass injected if it was overpredicted. The boundary condition for the lower cut point was between 1% and 80% of the peak maximum, and for the upper cut point between 20% and 99% of the peak maximum. The initial salt was between 5 and 300 mM, and the final salt between 100 and

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1200 mM for CEX and AEX. For HIC the initial salt was between 100 and 500 mM and the final salt concentration between 5 and 300 mM. The gradient length was varied between 1 and 10 CV. The computations were performed on a Dell Precision 5820 Tower XCTO having a 3.7G Intel Xeon processor of 3.7 GHz, 10 C, and a 8 GB Nvidia Quadro of 8GB. Multiple cores were used to execute the simulations most efficiently; however, the number of cores varied depending on the simulation.

2.4.3 | Artificial neural networks

The ANNs are developed using the Keras Module (version 2.4) of Tensorflow (version 2.3), both are open-source packages available in Python language. The ANN structure was defined using keras.models.Model and optimized using keras.optimizers.Adam, for which the learning rate was set to 0.001. Scaling of the data was done using the sklearn.preprocessing.MinMaxScaler module. The optimizer's loss function was set to "mean_squared_error," which is commonly applied for regression problems.

2.5 | Case study

For the case study, the product of interest, a monoclonal antibody (referred further as protein 1), and four impurities (referred further as proteins 2–5) were considered; data were taken from a previous study (Nfor et al., 2012). The protein names can be found in Appendix B. From the isotherm parameters it is expected that protein 1 elutes together with protein 5 in CEX, for AEX it is expected that protein 1 elutes together with protein 2, and partly with 3, and for HIC protein 1 will likely elute simultaneously with protein 4 and possibly partly with protein 5. Details of the isotherm and resin parameters can be found in Appendix B. The linear velocity was set to 150 cm/h. The initial concentration and amount of loaded product were varied for generating the data for creating ANNs. For the optimization part, the initial concentration of all proteins was set to 2 g/L with a loading factor of 2.0 CV.

The global and local objective were formulated as

$$\min f(x) = (100 - \text{yield}(x)) + 2 \times (100 - \text{purity}(x)) + \text{eluent consumption}(x), \quad (12)$$

$$\text{s. t. } h(x) = 0 \text{ (only applies to MM)}, \quad (13)$$

$$0 \le x \le 1, \quad (14)$$

where $f(x)$ is the objective function to be minimized, and the variables x can be operating or design parameters. All variables (x) were normalized between 0 and 1 for enhanced optimization purposes (Equation 14). Moreover, when using the MM, the equality equations $h(x)$ must be complied, which are mass balances and equilibrium relations (Equation 13). For optimizing each flowsheet, only operating variables (x) were considered, namely, the length of the gradient elution, initial and final salt concentration, lower and upper cut point. The performance measurements (e.g., yield, purity, buffer consumption) are calculated over the whole purification sequence. Purity was weighted twice as high, as purity is the most important factor for purifying biopharmaceuticals. By minimizing the buffer consumption, the costs, batch throughput and productivity are indirectly represented. The cost of lost feed is related to the yield. Subsequently, the selected best flowsheets and their most optimal conditions obtained after performing the global and minor local optimization were used as an input for the final local optimization.

3 | RESULTS & DISCUSSION

3.1 | Artificial neural networks

The ANNs were used as a meta-model during the global optimization to select the most promising flowsheet(s). Therefore, high accuracies of the ANNs are desired. Several steps were performed to build the ANNs, first high-quality data were generated, second the number of sample points was determined, and lastly the hyperparameters were optimized.

The accuracy of ANNs is relying on the quality of the data. The range of input variables is key, having a too broad range could lead to a poor accuracy on the data with lower values, while a too narrow range could lead to a biased optimization outcome and ANNs lacking flexibility. Details on the final range of parameters are given in Figure 2. The desired accuracy for the ANNs was an $R^2 > 0.90$ and RMSE < 0.04, as a trade-off has to be made between the number of sample points and the accuracy of the ANN. This RMSE value is normalized, transforming this value to the absolute value would give an error rate of about 15% on the mass of each protein, for the predicted volume and salt concentration of the product pool it was less than 15%. The mass input range was quite broad (4.81×10^{-8} –0.02 g), as both the mass and loading factor are input variables. We posited that an error rate of 15% would be acceptable for performing the flowsheet optimization, and with certainty identify the most optimal flowsheets while disregarding the less promising ones. Hence, to obtain this accuracy, the required number of sample points was evaluated for the product (Figure 3a,b for CEX). Ten ANNs were trained for each number of sample points, and the unseen test data were used for the boxplots. Increasing the number of sample points resulted in a higher R^2 and lower RMSE value, as expected. Using ANNs instead of MMs for the optimization would only be more efficient if less simulations are needed to generate the data than to run the optimization with the MM. Considering a flowsheet optimization of three chromatography modes, and assuming 15 flowsheets have to be evaluated 1000 times (Pirrung et al., 2017), this will result in a total of about 33,000 simulations with the MM. The total number of simulations can be derived by summing over the different types of flowsheets, namely three times one chromatography mode, six times two chromatography modes, and six times three

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

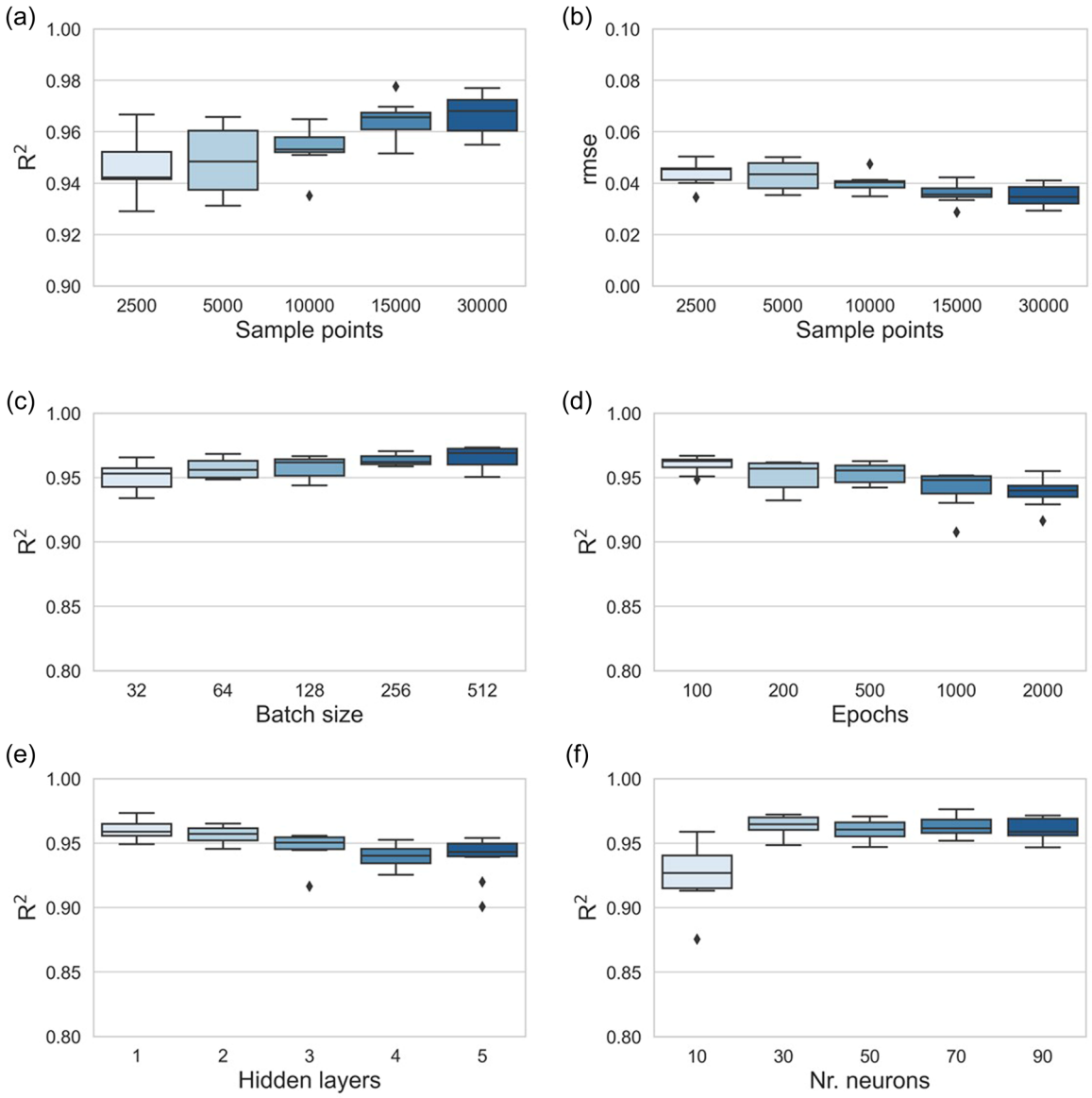


FIGURE 3 Boxplots a and b show the accuracy (left: R^2 and right: RMSE value) for different number of sample points for the CEX chromatography mode using the data of the product. Boxplots c–f show the effect of varying certain hyperparameters on the R^2 for protein 1 in CEX chromatography. The standard hyperparameters were three hidden layers each having 50 neurons, a batch size of 128 and epoch of 200, the number of sample points used was 10,000.

chromatography modes, which results in a total of 33,000 MM simulations. Consequently, a maximum of 10,000 simulations for generating the ANN data for each chromatography mode was desired. Based on this estimation and on the fact 10,000 sample points reached the desired accuracies, Figure 3a,b, we decided to continue with 10,000. The optimal ANN structure was identified by evaluating the effect of several hyperparameters (e.g., batch size, and number of epochs, hidden layers, and neurons) on the R^2 and RMSE value, Figure 3c–f for CEX. This overall evaluation for each chromatography mode and each protein can be found in Appendix C. The final hyperparameters were chosen based upon highest median for R^2 and lowest median for RMSE value. Moreover a small interquartile range (IQR) is desired, indicating less variance in accuracy. The used hyperparameters for each ANN are given in Table 1.

The quantitative evaluation of each ANN is shown in Table 2, most of the ANNs reached the desired values of $R^2 > 0.90$ and RMSE < 0.04. The generated data are focused on the product peak, and hence some proteins will never elute or be present in the product pool. As these output values were all very small, it is very hard to train the ANNs accurately, and so the R^2 remains low. However, the absolute RMSE is also very low ($<1 \times 10^{-5}$). As we know these proteins will never be present in the product pool, we could assume they would always be removed. The generated ANNs have sufficient predictive ability, as shown in Figure 4, for proteins 1, 4, and 5 during CEX for unseen test data. The data points are aligned close to the

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Quantitative evaluation for each chromatography mode and all proteins.

	Protein 1	Protein 2	Protein 3	Protein 4	Protein 5	Volume
CEX						
R^2	0.97	0.08	0.07	0.95	0.97	0.96
RMSE	0.032	0.153	0.176	0.027	0.034	0.029
AEX						
R^2	0.99	0.98	0.97	0.98	0.98	0.96
RMSE	0.022	0.027	0.026	0.009	0.009	0.035
HIC						
R^2	0.97	0.99	0.0	0.96	0.98	0.99
RMSE	0.037	0.025	0.5	0.041	0.028	0.034

Note: The calculations of each protein are based on the mass. The product pool volume is needed for connecting the unit operations and calculating the performance measurements during flowsheet optimization; therefore, this parameter is included.

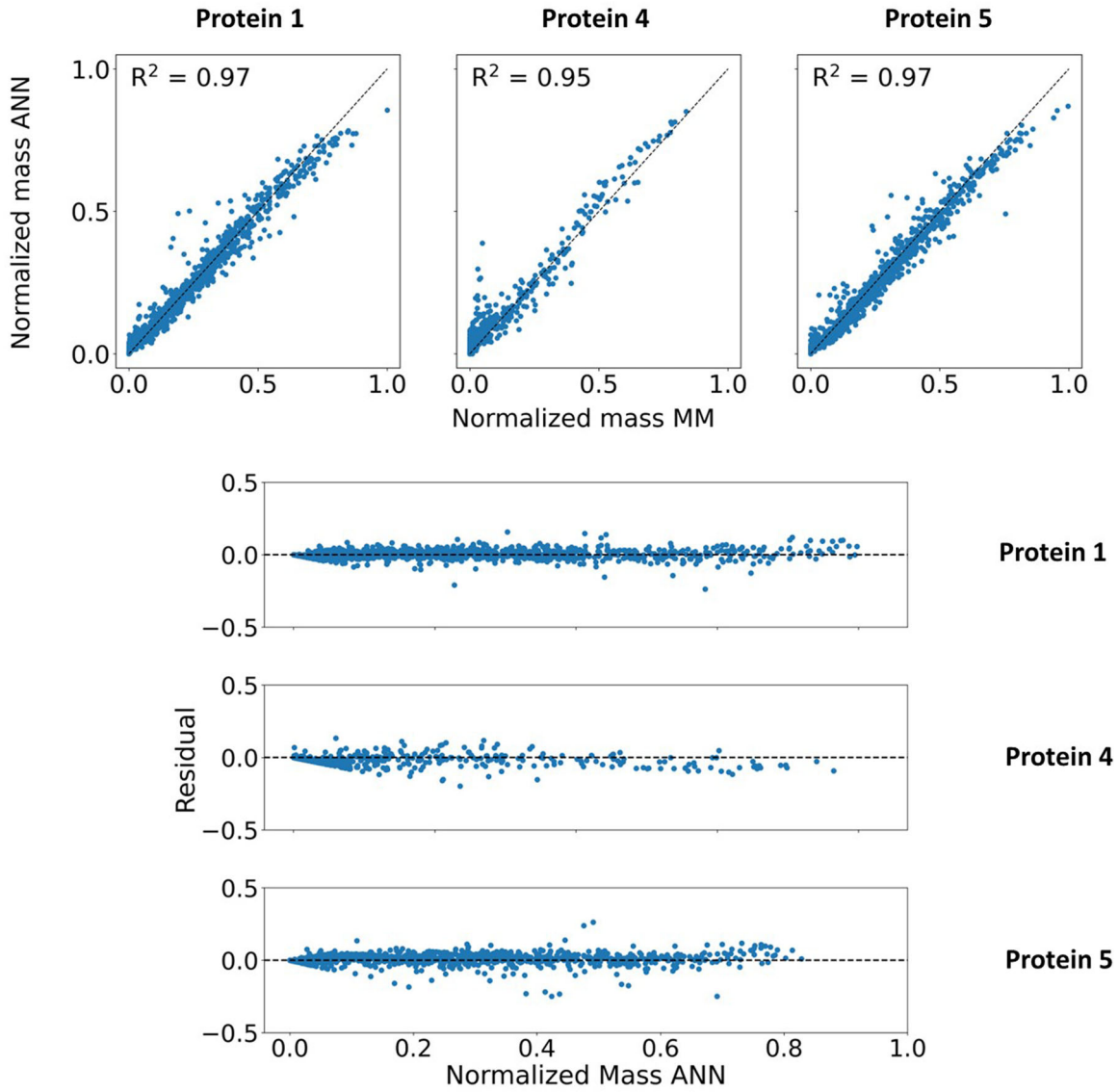


FIGURE 4 Upper figure: Prediction capabilities for the normalized artificial neural network (ANN) outcome of mass against the outcome of mechanistic model (MM). Lower figure: Residuals showing the difference between predicted mass values by the ANN and the MM. Both plots show unseen test data (1493 points) for the proteins 1, 4, and 5 for the CEX mode.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

diagonal, meaning the ANN is able to predict the outcome of the MM. The prediction capabilities for the other output variables and chromatography modes can be found in Appendix D. In addition to the R^2 and RMSE quantification, the residual plots assess the model's validity by evaluating the randomness in the residuals. In this case, all ANNs for the presented proteins show randomly scattered data points around the identity line, except for the proteins that were never present in the product pool, Figure 4 and Appendix D.

Contour plots for each chromatography mode were made to qualitatively evaluate the ANNs (Figure 5). These contour plots are used to evaluate if certain regions predicted by the ANNs are overpredicted or underpredicted, meaning the predicted ANN-values are higher, overprediction, or lower, underprediction, compared to the MM values. The ANN contour plots for both AEX and HIC are very similar to the MM contour plots (Figure 5 [2a,b and 3a,b]). However, all ANN contour plots show an over-prediction for a low lower-cut-point and high upper-cut-point compared to the MM results. While the ANN for CEX underpredicts the upper part of the lower cut point, hence when the cut point is closer to the end of the product peak (Figure 5 [1a]). The overprediction by ANNs is due to the standard deviation (SD) and results in an overprediction of the yield, for example, mass output divided by the mass injected.

3.2 | Flowsheet optimization

Optimizing a flowsheet is a multimodal optimization where multiple global optima could be present (Dominico & Parpinelli, 2021). In this optimization problem no information is known about the number of global optima, and the mathematical characteristics and gradient functions are also unknown. Therefore, we have chosen a stochastic and heuristic algorithm for the global and local optimization, respectively (e.g., Differential evolution and Nelder–Mead). This will enhance the likelihood of finding most of the global optima. To perform the flowsheet optimization within a reasonable amount of time, the number of function evaluations was defined for which the

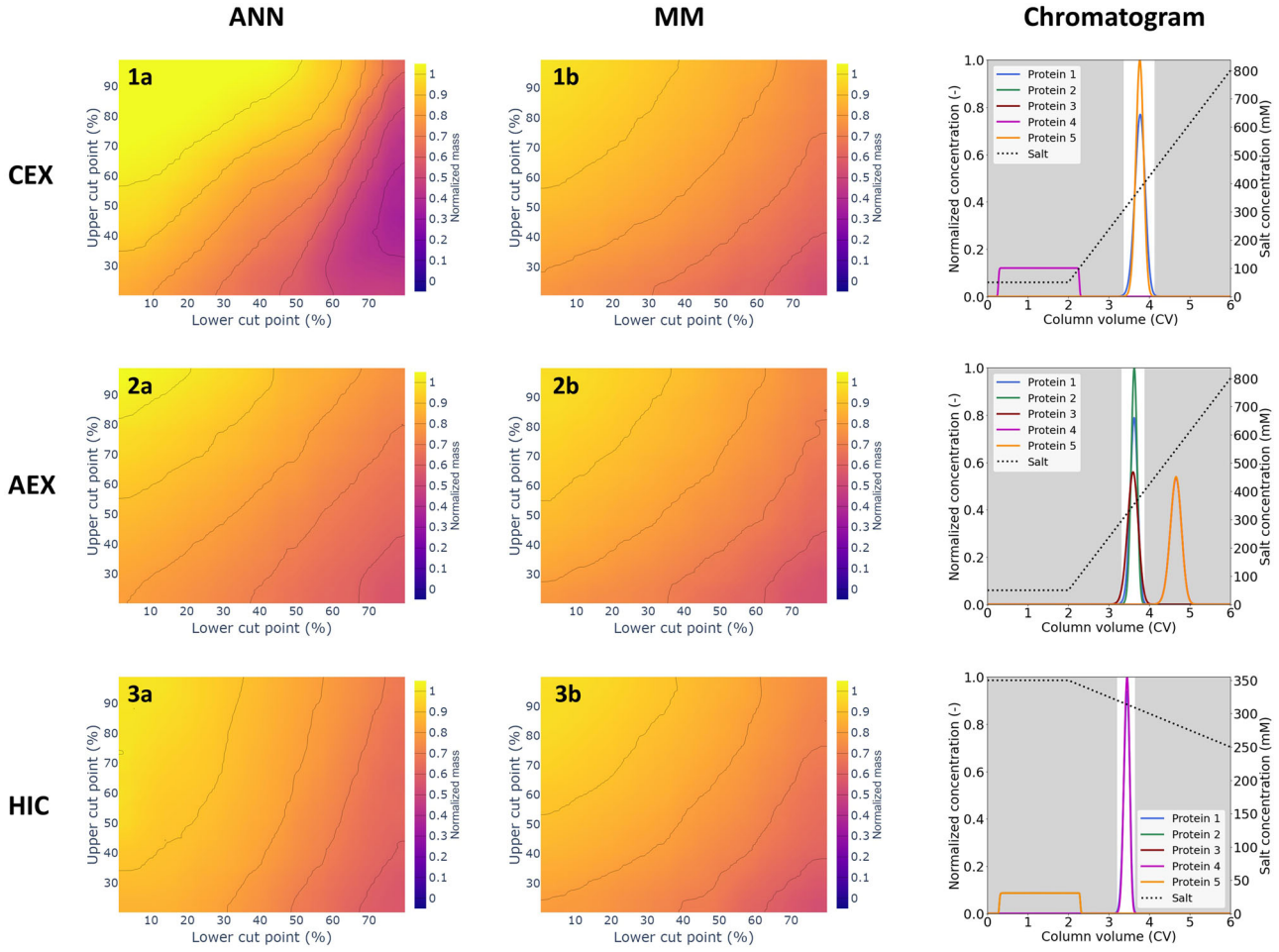


FIGURE 5 Qualitative analysis of the artificial neural networks (ANNs) compared to the true values of the mechanistic models (MMs) for each chromatography mode (CEX, AEX, and HIC). In the first column, the predicted results of the ANN for a varying range of cut points are shown, in the second column the outcome of the MM is shown. In the last column, the case study is shown including the initial and final salt concentration and gradient length. The loading factor was 2 CV and the inlet concentration 2.5 g/L or 0.00481 g in mass. The mass output is normalized to the mass injected (0.00481 g), also known as the yield.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

details can be found in Section 2.4: Numerical methods. The global optima are found when the function evaluations reach a plateau over several iterations, in this study a plateau is defined that the lowest 50 function evaluations have a maximum difference of 0.1 (Appendix E). This statement is needed, as often the number of maximum iterations is already reached before the relative and/or absolute tolerance are satisfied.

The optimization was performed for a superstructure of three chromatography modes and a maximum sequence of three unit operations, resulting in an evaluation of 15 flowsheets. The same operating conditions were used for the global optimization using either MMs or ANNs, which also applies to the optimization settings, except for the number of iterations that was increased when using ANNs. The flowsheets are evaluated by the WOP, which is based on the purity, yield, and buffer consumption (Section 2.1). The performance results of the global optimization using either MMs or ANNs is shown in Table 3.

Most of the performance outcomes between MM and ANN are comparable, as well as the calculated WOP. Only the sequences where CEX is the first unit operation failed to predict sufficient purity. As seen in Figure 5, the extreme cut points were over-predicted and underpredicted, so the ANN of CEX is not that accurate in this region. Consequently, the ANN is not able to find the global optima during the global optimization. However, it is remarkable that the ANN of CEX shows similar quantitative results as the other two chromatography modes (Table 2). The ANN of CEX can be improved to obtain the desired optimal accuracy, but that is not within the scope of this study, which aims to show the validity of the overall optimization approach. Most of the outcomes for sequences with three unit operations, predicted by the ANNs, imply the global optima were not found yet. The function evaluations show the plateau has not been reached, so more evaluations are needed to find the global optima (Appendix E). Due to the ANN's accuracy it is more difficult to find the global optima; therefore, more evaluations were allowed and even more would be needed for the sequences with three unit operations. The predictions for single unit operations show very similar results between MM and ANN; the same can be noticed for the two unit operation sequences starting with AEX or HIC. Only the predicted concentrations vary between MM and ANN; this could indicate different global optima were found. However, multiple global optima can be close to the same optimal objective value, while using different decision variables values. The same applies to the found global optima when only using MMs. A well-considered trade-off was made between number of sample points versus the ANN's accuracy. Even though different global optima were found between the MM and ANN, a confident decision can be made to select the most promising flowsheets and disregards the least promising ones. Ideally, a process employs a minimum number of unit operations. From the WOP results, we can draw the conclusion that a single unit operation is not sufficient to purify the product, but two unit operations can be sufficient. As two unit operations would be able to purify the product, the sequences with three unit operations can be disregarded. Although we considered the HIC sequences during the global optimization for showing the completeness of this approach, HIC is undesired to be the first unit operation as a buffer exchange step is needed before and after the process to increase and

TABLE 3 Optimization results after the global and minor local optimization using the MM and ANN.

Process option		Purity (%)		Yield (%)		Product concentration (g/L)		Buffer consumption (L/g)		WOP	
		MM	ANN	MM	ANN	MM	ANN	MM	ANN	MM	ANN
0	CEX	49.95	53.15	99.80	100.00	10.62	4.07	0.85	1.79	75	76
1	CEX–AEX	99.99	79.67	99.30	94.93	3.32	1.40	2.02	5.36	99	87
2	CEX–AEX–HIC	99.97	75.03	97.49	89.26	1.57	1.49	5.09	6.66	98	83
3	CEX–HIC	100.00	60.03	98.91	99.97	13.51	1.39	3.33	4.95	99	79
4	CEX–HIC–AEX	99.72	75.69	97.40	100.00	0.94	1.94	4.54	6.55	98	87
5	AEX	49.81	49.73	99.53	100.00	1.31	1.92	2.51	2.90	74	74
6	AEX–CEX	99.92	99.83	99.16	100.00	1.47	1.18	2.88	3.92	99	99
7	AEX–CEX–HIC	99.87	99.76	97.73	84.04	1.02	1.36	5.01	9.85	98	93
8	AEX–HIC	99.93	99.68	98.58	100.00	33.15	1.39	3.11	3.92	99	99
9	AEX–HIC–CEX	99.99	99.76	98.65	100.00	1.48	1.41	4.58	5.04	99	99
10	HIC	49.98	63.10	99.63	100.00	36.05	90.76	0.80	0.77	75	81
11	HIC–CEX	100.00	99.67	98.45	100.00	3.55	11.73	3.18	3.33	99	99
12	HIC–CEX–AEX	99.99	97.61	97.02	89.57	3.59	1.00	4.90	3.80	98	95
13	HIC–AEX	100.00	98.98	99.37	100.00	5.32	16.88	3.03	1.56	99	99
14	HIC–AEX –CEX	99.92	98.60	97.90	86.62	1.66	0.70	3.57	7.68	99	94

Abbreviations: ANN, artificial neural network; MM, mechanistic model; WOP, weighted overall performances.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

decrease the salt concentration. When using MMs, the found optimal sequences are 1, 3, 6, and 8 for a WOP > 85. For ANNs, the optimal sequences found are 1, 6, and 8. So, one sequence would be overlooked when only using ANNs for the global and minor local optimization. Nevertheless, most of the promising sequences to purify the product of interest are found with the ANNs. The identified sequences correspond to the results from Nfor et al. (2013). The performance results differ because other process conditions, objective, and variables were applied. Also, Pirrung et al. (2017) performed a similar study in which the optimal found sequence was CEX–HIC, in this study equal to sequence 3. Although different process conditions, objective, and variables were used, higher yield and purity values for all sequences were obtained in this study compared to the optimized results of Pirrung et al. (2017), when using MMs during the optimization. This can be assigned to different settings or to the fact that the global optima were not found yet.

As example, the global outcome of AEX–HIC for both MM and ANN was used as starting conditions to perform a final local optimization, for which similar results were found (Figure 6).

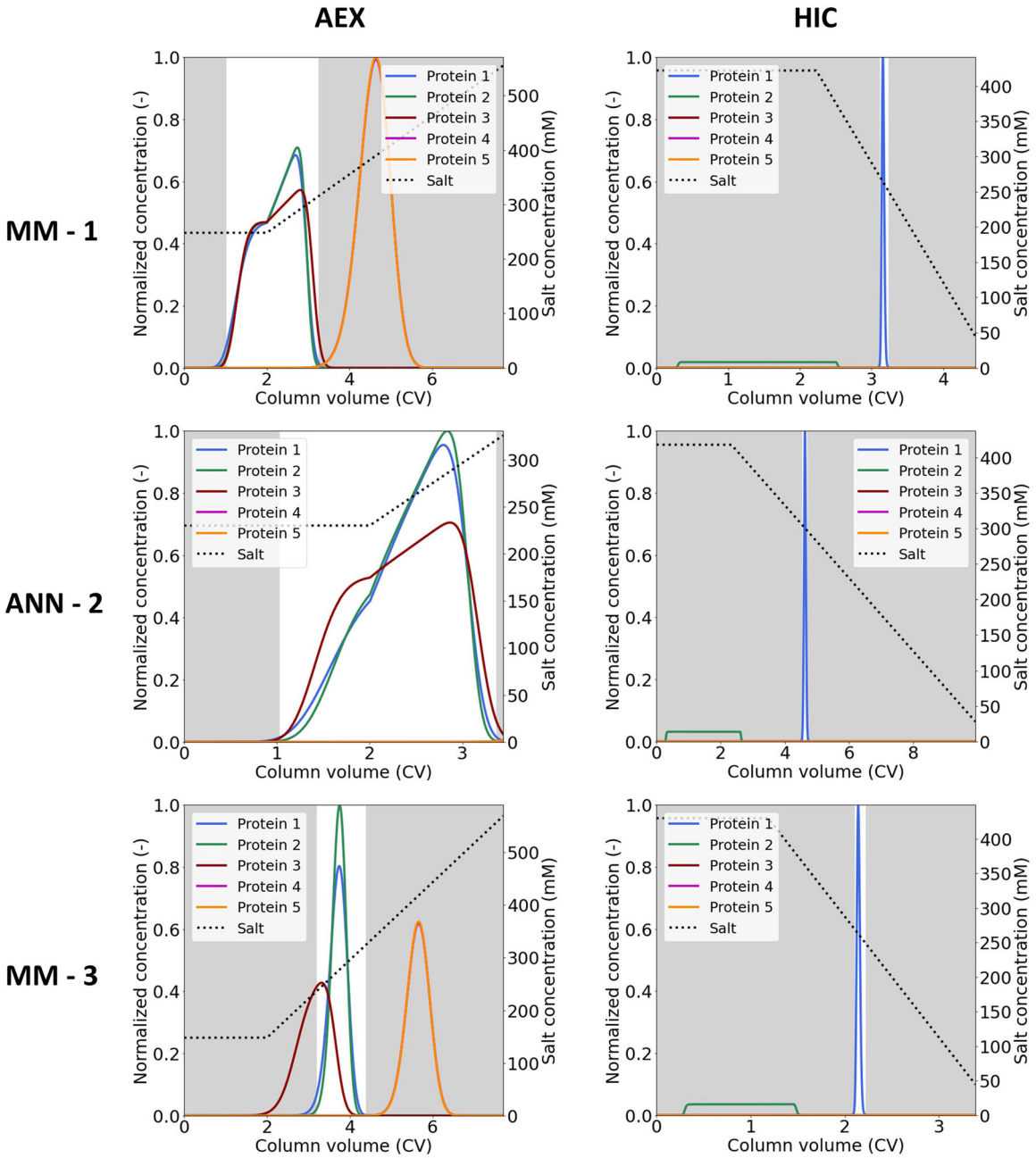


FIGURE 6 Outcome of the final local optimization using either the global outcome of the mechanistic model (MM) (first row) or the artificial neural network (ANN) (second row) as a starting condition. The third row (MM-3) shows the final local optimization outcome if the initial salt concentration is adjusted to a maximum of 150 mM. Protein 1 (mAb) is the target protein, to be separated from protein 2–5 (impurities).

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

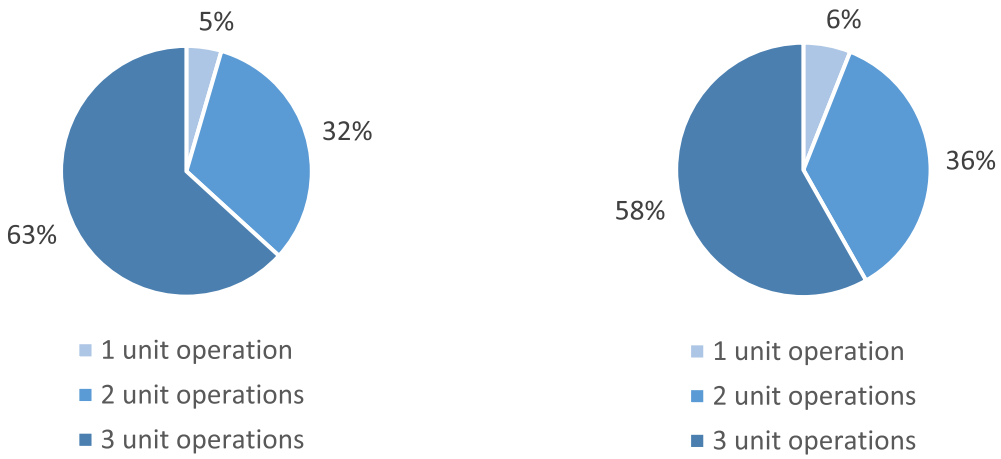
The range of the initial salt concentration varied between 5 and 300 mM. As a result, the predicted optimal conditions show an early elution of the product peak and a few impurities during the loading. This would be undesirable if more proteins or other impurities are present. However, the range of the initial salt concentration can be adjusted for both the global optimization or the final local optimization. An example is shown in Figure 6 (MM-3), where the maximum initial salt concentration for the final local optimization was adjusted to 150 mM. This also applies to the other input parameters. If the range is significantly different for the global optimization, it is recommended to train new ANNs to ensure accuracy.

Even though ANNs can be used for finding the optimal sequences during global optimization, they also should be beneficial for flowsheet optimization. An overview of time spent for global optimization using either MMs or ANNs is shown in Table 4. As expected for ANNs, about 97% of the total simulation time is spent on data generation, as MMs are used for this task. The data generation also includes the training of ANNs; however, the time required to complete this task is minimal. Much more optimization evaluations can be performed using ANNs, but also more evaluations are required to find sufficient results. The simulation time for the different length of sequences is similar for both the MM and ANN (Table 4). Overall, ANNs are twice as fast compared to MMs for this flowsheet optimization. To make a fair comparison, optimal parallelization was excluded for this study; however, both approaches would benefit from parallelization to decrease the overall simulation time. The minor local optimization, included within the overall global optimization time, took about 15 h for the mechanistic model and 0.08 h for the ANNs, which did not have a significant influence on the overall time. As both frameworks use the MM for the final local optimization, the duration was also similar for both frameworks.

This approach becomes especially advantageous when evaluating larger superstructures, involving either more unit operation modes and/or larger sequence lengths. For example, considering five different resins in a maximum sequence of three unit operations, 85 flowsheets have to be evaluated. This will take approximately 95 days when using MMs. For ANNs, the optimization will only take 1.3 days, and generating data for five different resins will take about 14 days. Hence, using ANNs for this larger superstructure will be 6.4 times faster than using MMs. For a process design where more proteins are considered, it is expected that both approaches would need about similar extra simulations time.

TABLE 4 Comparison of time spent for the global optimization when using MMs or ANNs.

Mechanistic model		Artificial Neural Network	
		Data generation:	195 hrs.
Global optimization:	399 hrs.	Global optimization:	5.6 hrs.
Local optimization:	2.9 hrs.	Local optimization:	6.4 hrs.
Total time:	401.9 hrs.	Total time:	207 hrs.
Total global evaluations:	15500	Total global evaluations:	52800
Total local evaluations:	128	Total local evaluations:	308



Note: The data generation for the ANNs includes simulations of 10,000 sample points for each chromatography mode and the training of the ANNs.
Abbreviations: ANN, artificial neural network; MM, mechanistic model.

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | CONCLUSIONS

In this study, we have compared two optimization frameworks for purifying biopharmaceuticals, by either employing MMs or ANNs for global optimization. The global optimization outcome was used to pre-select the most optimal process sequences, which subsequently were optimized locally. Three types of chromatography were considered during the optimization. First, we built the ANNs for each chromatography mode, most of them reached an accuracy of $R^2 > 0.95$ and RMSE < 0.04. Next, we performed a flowsheet optimization for a superstructure of 15 flowsheets. Our results proved that ANNs can be used during global optimization to make a pre-selection for the most optimal process-sequences according to certain objectives and constraints. The final local optimization results were comparable when using either the global outcome of the MMs or the ANNs as starting condition. The overall computation of the global optimization when using MMs took about 400 h, while using ANNs took about half of the time so 200 h.

To make ANNs more accurate, the data acquisition has to be tuned, for example, narrowing the design space of the input parameters. Though, by incorporating more knowledge, ANNs will also become more biased and less flexible. Another approach is to develop several ANNs for specific regions of the input parameters. In this study, we chose to make one ANN to reduce complexity, and a broader range of input parameters to remain flexible and less biased. Though, at the expense of accuracy.

This study represents a step toward a new model-based approach for developing biopharmaceutical purification processes. This is especially important for early conceptual process design, when a limited amount of sample material is available and little is known about the sample's purification process. This study provides a generic way to develop ANNs for downstream processes and shows the usefulness of ANNs in accelerating flowsheet optimizations. In fact, for this case-study, using ANNs during flowsheet optimization reduced the computational time by 50% compared to using only MMs. For larger superstructures ANNs could even be an order of magnitude times faster than shown for this superstructure consisting of 15 flowsheets.

AUTHOR CONTRIBUTIONS

Daphne Keulen: Conceptualization; methodology; investigation; software; simulations; data analysis; visualization; writing—original draft; writing—review and editing. **Erik van der Hagen**: Software; methodology; simulations; writing—review and editing. **Geoffroy Geldhof**: Supervision; writing—review & editing. **Olivier Le Bussy**: Supervision; writing—review & editing. **Martin Pabst**: Conceptualization; writing—review & editing. **Marcel Ottens**: Conceptualization; supervision; interpretation; writing—review & editing; funding acquisition.

ACKNOWLEDGMENTS

This study was funded by GlaxoSmithKline Biologicals S.A. under cooperative research and development agreement between GlaxoSmithKline Biologicals S.A. (Belgium) and the Technical University of Delft (The Netherlands). The authors thank the colleagues from GSK and Technical University of Delft for their valuable input.

CONFLICTS OF INTEREST STATEMENT

All authors have declared the following interests: Geoffroy Geldhof and Olivier Le Bussy are employees of the GSK group of companies. The remaining authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article

ORCID

Daphne Keulen http://orcid.org/0000-0001-8086-333X
Martin Pabst http://orcid.org/0000-0001-9897-0723

REFERENCES

Bhambure, R., & Rathore, A. S. (2013). Chromatography process development in the quality by design paradigm I: Establishing a high-throughput process development platform as a tool for estimating "characterization space" for an ion exchange chromatography step. *Biotechnology Progress*, *29*(2), 403–414. https://doi.org/10.1002/btpr.1705

Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., & Ierapetritou, M. (2020). Digital twins in pharmaceutical and biopharmaceutical manufacturing: A literature review. *Processes*, *8*(9), 1088. https://doi.org/10.3390/pr8091088

Dominico, G., & Parpinelli, R. S. (2021). Multiple global optima location using differential evolution, clustering, and local search. *Applied Soft Computing*, *108*, 107448. https://doi.org/10.1016/j.asoc.2021.107448

FDA. (2004). *PAT guidance for industry—A framework for innovative pharmaceutical development, manufacturing and quality assurance*. www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance

Felinger, A., & Guiochon, G. (2004). Comparison of the kinetic models of linear chromatography. *Chromatographia*, *60*(1), S175–S180. https://doi.org/10.1365/s10337-004-0288-7

Fellner, M., Delgado, A., & Becker, T. (2003). Functional nodes in dynamic neural networks for bioprocess modelling. *Bioprocess and Biosystems Engineering*, *25*(5), 263–270. https://doi.org/10.1007/s00449-002-0297-6

Huuk, T. C., Hahn, T., Osberghaus, A., & Hubbuch, J. (2014). Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Separation and Purification Technology*, *136*, 207–222. https://doi.org/10.1016/j.seppur.2014.09.012

ICH. (2009). *ICH harmonised tripartite guideline: Pharmaceutical development Q8 (R2)* ICH. https://www.ema.europa.eu/en/ich-q8-r2-pharmaceutical-development-scientific-guideline

Kawajiri, Y. (2021). Model-based optimization strategies for chromatographic processes: A review. *Adsorption*, *27*(1), 1–26. https://doi.org/10.1007/s10450-020-00251-2

Keulen, D., Geldhof, G., Bussy, O. L., Pabst, M., & Ottens, M. (2022). Recent advances to accelerate purification process development: A review with a focus on vaccines. *Journal of Chromatography A*, *1676*, 463195. https://doi.org/10.1016/j.chroma.2022.463195

Lin, D.-Q., Zhang, Q.-L., & Yao, S.-J. (2021). Model-assisted approaches for continuous chromatography: Current situation and challenges. *Journal of Chromatography A*, *1637*, 461855. https://doi.org/10.1016/j.chroma.2020.461855

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Łącki, K. M. (2018). Chapter 16—Introduction to preparative protein chromatography. In G. Jagschies, E. Lindskog, K. Łącki, & P. Galliher (Eds.), *Biopharmaceutical Processing* (pp. 319–366). Elsevier. https://doi.org/10.1016/B978-0-08-100623-8.00016-5

Madden, J. E., Avdalovic, N., Haddad, P. R., & Havel, J. (2001). Prediction of retention times for anions in linear gradient elution ion chromatography with hydroxide eluents using artificial neural networks. *Journal of Chromatography A*, *910*(1), 173–179. https://doi.org/10.1016/S0021-9673(00)01185-7

Müller, A. C., & Guido, S. (2017). *Introduction to machine learning with python: A guide for data scientists* O'Reilly Media.

Nagrath, D., Messac, A., Bequette, B. W., & Cramer, S. M. (2004). A hybrid model framework for the optimization of preparative chromatographic processes. *Biotechnology Progress*, *20*(1), 162–178. https://doi.org/10.1021/bp034026g

Narayanan, H., Seidler, T., Luna, M. F., Sokolov, M., Morbidelli, M., & Butté, A. (2021). Hybrid models for the simulation and prediction of chromatographic processes for protein capture. *Journal of Chromatography A*, *1650*, 462248. https://doi.org/10.1016/j.chroma.2021.462248

Nfor, B. K., Ahamed, T., van Dedem, G. W. K., Verhaert, P. D. E. M., van der Wielen, L. A. M., Eppink, M. H. M., van de Sandt, E. J. A. X., & Ottens, M. (2013). Model-based rational methodology for protein purification process synthesis. *Chemical Engineering Science*, *89*, 185–195. https://doi.org/10.1016/j.ces.2012.11.034

Nfor, B. K., Ahamed, T., Pinkse, M. W. H., van der Wielen, L. A. M., Verhaert, P. D. E. M., van Dedem, G. W. K., Eppink, M. H. M., van de Sandt, E. J. A. X., & Ottens, M. (2012). Multi-dimensional fractionation and characterization of crude protein mixtures: Toward establishment of a database of protein purification process development parameters. *Biotechnology and Bioengineering*, *109*(12), 3070–3083. https://doi.org/10.1002/bit.24576

Nfor, B. K., Noverraz, M., Chilamkurthi, S., Verhaert, P. D. E. M., van der Wielen, L. A. M., & Ottens, M. (2010). High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents. *Journal of Chromatography A*, *1217*(44), 6829–6850. https://doi.org/10.1016/j.chroma.2010.07.069

Nfor, B. K., Zuluaga, D. S., Verheijen, P. J. T., Verhaert, P. D. E. M., van der Wielen, L. A. M., & Ottens, M. (2011). Model-based rational strategy for chromatographic resin selection. *Biotechnology Progress*, *27*(6), 1629–1643. https://doi.org/10.1002/btpr.691

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv*. https://doi.org/10.48550/arXiv.1811.03378

Petzold, L. (1983). Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing*, *4*(1), 136–148. https://doi.org/10.1137/0904010

Pirrung, S. M., Berends, C., Backx, A. H., van Beckhoven, R. F. W. C., Eppink, M. H. M., & Ottens, M. (2019). Model-based optimization of integrated purification sequences for biopharmaceuticals. *Chemical Engineering Science: X*, *3*, 100025. https://doi.org/10.1016/j.cesx.2019.100025

Pirrung, S. M., van der Wielen, L. A. M., van Beckhoven, R. F. W. C., van de Sandt, E. J. A. X., Eppink, M. H. M., & Ottens, M. (2017). Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. *Biotechnology Progress*, *33*(3), 696–707. https://doi.org/10.1002/btpr.2435

Portela, R. M. C., Varsakelis, C., Richelle, A., Giannelos, N., Pence, J., Dessoy, S., & von Stosch, M. (2020). When is an in silico representation a digital twin? A biopharmaceutical industry approach to the digital twin concept, *Digital Twins* (pp. 35–55). Springer. https://doi.org/10.1007/10-2020-138

Rathore, A. S., Nikita, S., Thakur, G., & Mishra, S. (2023). Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. *Trends in Biotechnology*, *41*(4), 497. https://doi.org/10.1016/j.tibtech.2022.08.007

Reinhardt, I. C., Oliveira, D. J. C., & Ring, D. D. T. (2020). Current perspectives on the development of industry 4.0 in the pharmaceutical sector. *Journal of Industrial Information Integration*, *18*, 100131. https://doi.org/10.1016/j.jii.2020.100131

Ruthven, D. M. (1984). *Principles of adsorption and adsorption processes*. John Wiley & Sons.

Schmölder, J., & Kaspereit, M. (2020). A modular framework for the modelling and optimization of advanced chromatographic processes. *Processes*, *8*(1), 65. https://doi.org/10.3390/pr8010065

Silva, F., Resende, D., Amorim, M., & Borges, M. (2020). A field study on the impacts of implementing concepts and elements of industry 4.0 in the biopharmaceutical sector. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(4), 175. https://doi.org/10.3390/joitmc6040175

Stamatis, C., Goldrick, S., Gruber, D., Turner, R., Titchener-Hooker, N. J., & Farid, S. S. (2019). High throughput process development workflow with advanced decision-support for antibody purification. *Journal of Chromatography A*, *1596*, 104–116. https://doi.org/10.1016/j.chroma.2019.03.005

von Stosch, M., Oliveira, R., Peres, J., & de Azevedo, S. F. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101. https://doi.org/10.1016/j.compchemeng.2013.08.008

von Stosch, M., Portela, R. M., & Varsakelis, C. (2021). A roadmap to AI-driven in silico process development: Bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering*, *33*, 100692. https://doi.org/10.1016/j.coche.2021.100692

Yeomans, H., & Grossmann, I. E. (1999). A systematic modeling framework of superstructure optimization in process synthesis. *Computers & Chemical Engineering*, *23*(6), 709–731. https://doi.org/10.1016/S0098-1354(99)00003-4

Yu, L. X. (2008). Pharmaceutical quality by design: Product and process development, understanding, and control. *Pharmaceutical Research*, *25*(4), 781–791. https://doi.org/10.1007/s11095-007-9511-1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Keulen, D., Hagen, E. v. d., Geldhof, G., Le Bussy, O., Pabst, M., & Ottens, M. (2023). Using artificial neural networks to accelerate flowsheet optimization for downstream process development. *Biotechnology and Bioengineering*, 1–14. https://doi.org/10.1002/bit.28454

10970290, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.28454 by Tu Delft, Wiley Online Library on [19/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License