Computational modeling and optimization of biopharmaceutical downstream processes

Keulen, D.

DOI
10.4233/uuid:f55e8d73-d9c5-4e38-bb8b-43a87301ef82

Publication date
2024

Document Version
Final published version

Citation (APA)
Keulen, D. (2024). *Computational modeling and optimization of biopharmaceutical downstream processes*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:f55e8d73-d9c5-4e38-bb8b-43a87301ef82

Computational modeling and optimization of biopharmaceutical downstream processes

Computational modeling and optimization of biopharmaceutical downstream processes

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 22 May 2024 at 12:30 o'clock

by

Daphne KEULEN

Master of Science in Life Science and Technology,
Delft University of Technology, the Netherlands
born in Geldrop, the Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. M. Ottens	Delft University of Technology, promotor
dr. M. Pabst	Delf University of Technology, copromotor

Independent members:

Prof. dr. M.R. Aires-Barros	Instituto Superior Técnico, Portugal
Prof. dr. B. Nilsson	Lund University, Sweden
Prof. dr. ir. C.A. Ramirez Ramirez	Delft University of Technology
Prof. dr. ing. M.H.M. Eppink	Delft University of Technology / Byondis
Prof. dr. ir. L.A.M. van der Wielen	Delft University of Technology / University of Limerick

This research 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).

Printed by Proefschriftspecialist

Cover illustration by Vera van Maaren

Dedicated to my father

Table of contents

Summary

Vaccination is a strong and effective way to prevent spreading of infectious diseases and promotes global health. In the future, the importance of vaccines is expected only to increase, driven by factors such as increased international traveling, higher healthcare expenditures, and a growing population. To meet the growing demands, it is necessary to shorten process development timelines for the production of new vaccines without compromising on safety, efficacy, consistency, and stability of the product. Therefore, it is necessary to advance process development approaches of vaccines to respond quickly and in event of an emerging infectious disease. The work in this thesis employed mathematical modeling and simulation techniques to accelerate this process development. The developed methods are particularly valuable for early phase process development, aiming to enhance process knowledge and minimize the consumption of valuable resources and material. While the project has a focus on vaccine production processes, the modeling tools and methods developed are also applicable to other (bio)pharmaceutical processes. In **Chapter 2**, we discussed the present and future process development approaches in (bio)pharmaceutical purification with an emphasis on vaccines. The primary needs are to establish standardized processes and to improve understanding of both production processes and host cell impurities. Modeling, when combined with high throughput experimentation, can play a crucial role in achieving these goals.

We used mechanistic modeling (MM) to mathematically describe the physical phenomena occurring in a real process. As chromatography can attain very high product purities, this is one of the main purification techniques for vaccines and therefore one of our central focuses throughout this thesis. Identifying an optimal purification process early in the development phase is advantageous considering costs, quality, and development time. Flowsheet optimization evaluates all potential process sequences *in silico* and therefore enables selecting the most optimal process(es) in the early phase of the process development. An optimization software was developed to perform such complex flowsheet optimization, which is described in **Chapter 3**. However, during flowsheet optimization, chromatographic mechanistic modeling can be time consuming and speed limiting, and therefore artificial neural networks (ANNs) were developed. Artificial neural networks functioned as surrogate models of the mechanistic model, with the goal of reducing overall computational time while still identifying the most optimal sequence(s). In this chapter, we compared the utilization of both artificial neural networks and mechanistic modeling during flowsheet optimization in terms of outcomes and computational time. Our results demonstrated that artificial neural networks can be used during global optimization to pre-select the most optimal process sequences

based on defined objectives and constraints. The overall computational time, including data generation and artificial neural networks training, is reduced by 50% when using artificial neural networks.

Apart from the modeling technique as described in Chapter 3, the optimization strategy itself appeared to be just as important in terms of outcome, complexity, and time-efficiency (**Chapter 4**). In this chapter, we compared three optimization strategies along with each strategy being optimized by both mechanistic modeling and artificial neural networks. Moreover, an optional buffer exchange was included between the chromatography steps, which increased the complexity of the flowsheet optimization. The three optimization strategies (e.g., simultaneous, top-to-bottom, and superstructure decomposition) differed in their approach to optimize the sequence of unit operations, whether all at once, in parts, or individually. The superstructure decomposition strategy with mechanistic modeling was found to be the most time efficient method, a complete flowsheet optimization considering 39 flowsheets was performed around a day using a state-of-the-art computer workstation.

Adsorption isotherm parameters are essential input parameters for mechanistic modeling. The determination of these parameters is typically done experimentally, and remains a bottleneck for mechanistic modeling of adsorption in process development. An alternative *in silico* method is quantitative structure property relationship (QSPR), which can predict retention times or specific adsorption isotherms based on the structure of individual proteins by correlating physiochemical properties. In **Chapter 5**, we developed a multiscale modeling approach by integrating quantitative structure property relationship with mechanistic modeling. The quantitative structure property relationship-based adsorption isotherm parameters were used in the mechanistic model. The validated mechanistic model showed a strong agreement with the experimental data, as only 0.2% difference between the retention peak values was observed, relative to the salt gradient length. Subsequently, the validated mechanistic model was employed to optimize a chromatographic capture step.

This work highlights the value of modeling approaches in process development. The application of different modeling techniques and optimization strategies during flowsheet optimization can guide in finding a suitable approach for a given case study. Furthermore, the multiscale modeling approach demonstrated its potential for industrial applications, allowing to find an optimal process without doing any initial experiments.

Samenvatting

Vaccineren is een effectieve manier om de verspreiding van infectieziekten te voorkomen en bevordert daarmee wereldwijd de gezondheid. Het belang van vaccins zal naar verwachting in de toekomst alleen maar toenemen als gevolg van toegenomen internationale reizen, hogere gezondheidskosten en een groeiende wereldpopulatie. Om snel te kunnen handelen en aan een toenemende vraag te kunnen voldoen als er een infectieziekte opkomt, is het noodzakelijk om de procesontwikkelingstijd voor de productie van nieuwe vaccins te verkorten zonder daarbij concessies te doen aan veiligheid, werkzaamheid, consistentie, en stabiliteit van het product. Het werk verricht in dit proefschrift omvat wiskundige modeleer- en simulatietechnieken om zo de procesontwikkeling voor de productie van vaccins te kunnen versnellen. De ontwikkelde methoden zijn met name waardevol in de beginfase van de procesontwikkeling om proceskennis te vergroten en tegelijkertijd het gebruik van kostbare middelen en materialen te minimaliseren. Alhoewel het project gericht is op het productieproces van vaccins, kunnen de ontwikkelde modellen en methoden ook worden toegepast op andere (bio)farmaceutische productieprocessen. In **Hoofdstuk 2** worden de huidige en mogelijk toekomstige benaderingen voor proces ontwikkeling in (bio)farmaceutische productzuivering besproken, met een nadruk op vaccins. Hieruit blijkt dat het essentieel is om een standaardproces te ontwikkelen en de kennis met betrekking tot het productieproces en de onzuiverheden van de gastheercel te verbeteren. Wiskundige modellen kunnen hierbij een cruciale rol spelen wanneer deze worden gecombineerd met geautomatiseerde experimenten.

Mechanistische modellen worden gebruikt om wiskundig de fysische verschijnselen te kunnen beschrijven die plaatsvinden tijdens het echte proces. Aangezien met chromatografie zeer hoge product zuiverheden bereikt kunnen worden, is dit een van de belangrijkste zuiveringstechnieken voor vaccins en daarmee een van de hoofdthema's in dit proefschrift. Het is bevorderlijk om in een vroeg stadium van de procesontwikkeling een optimaal zuiveringsproces vast te stellen, met betrekking tot kosten, kwaliteit en ontwikkelingstijd. Met *flowsheet*-optimalisatie worden alle potentiële volgordes van processtappen *in silico* geëvalueerd. Dit maakt het mogelijk om in de beginfase van de procesontwikkeling de meest optimale proces(sen) te selecteren. Voor het uitvoeren van dergelijke complexe *flowsheet*-optimalisaties is een optimalisatie software ontwikkeld, zoals beschreven in **Hoofdstuk 3**. Tijdens *flowsheet*-optimalisaties kunnen chromatografische mechanistische modellen echter tijdrovend en beperkend in rekensnelheid zijn, daarom zijn er kunstmatige neurale netwerken (*Artificial Neural Networks* - ANNs) ontwikkeld.

Kunstmatige neurale netwerken dienen als een vereenvoudigd model van het mechanistische model, met als doel de totale rekentijd te verminderen, terwijl nog steeds het meest optimale proces kan worden geselecteerd. De toepassing van zowel kunstmatige neurale netwerken als mechanistische modellen wordt vergeleken tijdens de *flowsheet*-optimalisatie, voor zowel de behaalde resultaten als de benodigde rekentijd. De resultaten tonen aan dat kunstmatige neurale netwerken gebruikt kunnen worden om de meest optimale processen vooraf te selecteren tijdens de globale optimalisatie op basis van vastgestelde doelstellingen en randvoorwaarden. De totale rekentijd, inclusief het genereren van de data en het trainen van de kunstmatige neurale netwerken, vermindert met 50% bij het gebruik van kunstmatige neurale netwerken.

Naast de verschillende modelleer technieken zoals beschreven in Hoofdstuk 3, blijkt de optimalisatiestrategie zelf net zo belangrijk te zijn voor wat betreft het resultaat, de complexiteit en de rekentijd in *flowsheet*-optimalisatie. In **Hoofdstuk 4** worden de optimalisatiestrategieën vergeleken, waarbij zowel mechanistische modellen als kunstmatige neurale netwerken zijn gebruikt. Daarnaast is een optioneel filtratieproces toegevoegd om buffers te verwisselen tussen de chromatografie stappen, waardoor de complexiteit van de *flowsheet*-optimalisatie toeneemt. De drie gebruikte optimalisatiestrategieën (e.g., simultaan, top-to-bottom en superstructuurdecompositie) verschillen in hun benadering om de volgorde van de processtappen te optimaliseren: allemaal tegelijk; opgedeeld; of individueel. De superstructuurdecompositie' strategie uitgevoerd met mechanistische modellen blijkt de meest rekentijds-efficiënte methode te zijn. Een volledige *flowsheet*-optimalisatie van 39 flowsheets duurt ongeveer een dag, uitgevoerd met een geavanceerde state-of-the-art computer.

Adsorptie-isothermparameters zijn essentiële ingrediënten voor een mechanistisch chromatografie model. De bepaling van deze parameters gebeurt doorgaans experimenteel en dit blijft daardoor een obstakel voor de grootschalige toepassing van mechanistische modellen voor adsorptie in procesontwikkeling. Een alternatieve *in silico* methode is om gebruik te maken van kwantitatieve structuur-eigenschapsrelaties (*Quantitative Structure Property Relationships* - QSPR). Deze modellen kunnen de chromatografische retentietijd danwel specifieke adsorptie-isothermparameters voorspellen op basis van de individuele eiwitstructuur door fysisch-chemische eigenschappen te correleren. In **Hoofdstuk 5** hebben we de kwantitatieve structuur-eigenschapsrelaties geïntegreerd met mechanistische chromatografie modellen om zo een modelleer aanpak op meervoudig niveau te ontwikkelen. De adsorptie-isothermparameters verkregen via kwantitatieve structuur-eigenschapsrelaties, worden nu gebruikt in het mechanistisch model. Het gevalideerde mechanistische model komt zeer goed overeen met de experimentele data, met slechts 0.2% verschil tussen de

retentiepieken relatief t.o.v. de lengte van de zoutgradiënt voor elutie. Uiteindelijk is dit gevalideerde model gebruikt om een chromatografische zuiveringstap te optimaliseren.

Dit werk benadrukt de toegevoegde waarde van mathematisch modeleren in procesontwikkeling. De toepassing van verschillende modelleertechnieken en optimalisatiestrategieën tijdens *flowsheet*-optimalisatie kan als leidraad dienen bij het vinden van een geschikte aanpak voor een bepaalde casus. Tevens toont de integrale modelleeraanpak, waarbij kwantitatieve structuur-eigenschapsrelaties worden gecombineerd met mechanistische modellen, haar potentieel voor industriële toepassingen. Hierdoor is het uiteindelijk mogelijk om een optimaal proces te ontwerpen zonder enige initiële experimenten uit te voeren.

Chapter 1

General introduction and Thesis outline

1.1. Background and aim

Communicable diseases, commonly referred as infectious diseases, account for approximately a quarter of all global deaths, of which 90% occur in low- and middle-income countries [1, 2]. In these countries, communicable diseases were responsible for more than half of the mortality among children and adolescents (e.g., 0 – 24 years), compared to only 5.6% of deaths in high-income countries [1]. Vaccination is pivotal in preventing and controlling the spread of infectious diseases. It significantly decreases the morbidity and mortality rates associated with vaccine-preventable diseases [2]. For instance, vaccination played a crucial role in completely eradicating smallpox worldwide. In recent years, the world has experienced several epidemics and pandemic, including the Zika virus, Ebola and COVID-19, which have led to a worldwide increased awareness about the value of vaccines. In addition to improved public health, vaccination also lowers the healthcare expenses, fostering economic growth, ensuring travel safety, and extending life expectancy [3, 4]. Vaccines are a subgroup of biopharmaceuticals, which are medications derived from or containing components of living organisms. The production processes for various biopharmaceuticals share broad similarities. Consequently, the methods developed in this thesis can be applied to other biopharmaceutical productions.

The global revenues in the vaccine market are expected to reach almost 82 billion US dollars in 2023, which is a significant increase compared to 26 billion US dollars in 2016 [2, 5]. In 2021, the reported value of the global biopharmaceutical market was approximately 343 billion US dollars, indicating that the market share of vaccines is about 24% [6]. COVID-19 had a substantial impact in the biopharmaceutical industry, with two of the top-selling biopharmaceuticals in 2021 being Comirnaty by Pfizer & BioNTech and Spikevax by Moderna. Together, these vaccines generate a cumulative revenue of 54.5 billion US dollars [6]. In the coming years, it is expected that the revenues from COVID-19 vaccines will decline, while other vaccines are expected to exhibit a steady upward trend, as illustrated in Figure 1.1 [5]. The leader companies in vaccines, excluding COVID-19 vaccines, are GSK, Merck & Co, Sanofi, and Pfizer, with their respective market shares in percentages shown in Figure 1.1 [2, 5]. UNICEF plays a crucial role in delivering vaccines to children and young adults in need worldwide [7]. One of three largest suppliers to UNICEF is the Serum Institute of India (SSI), a global leader in vaccine production, providing over 1.3 billion doses annually [2].

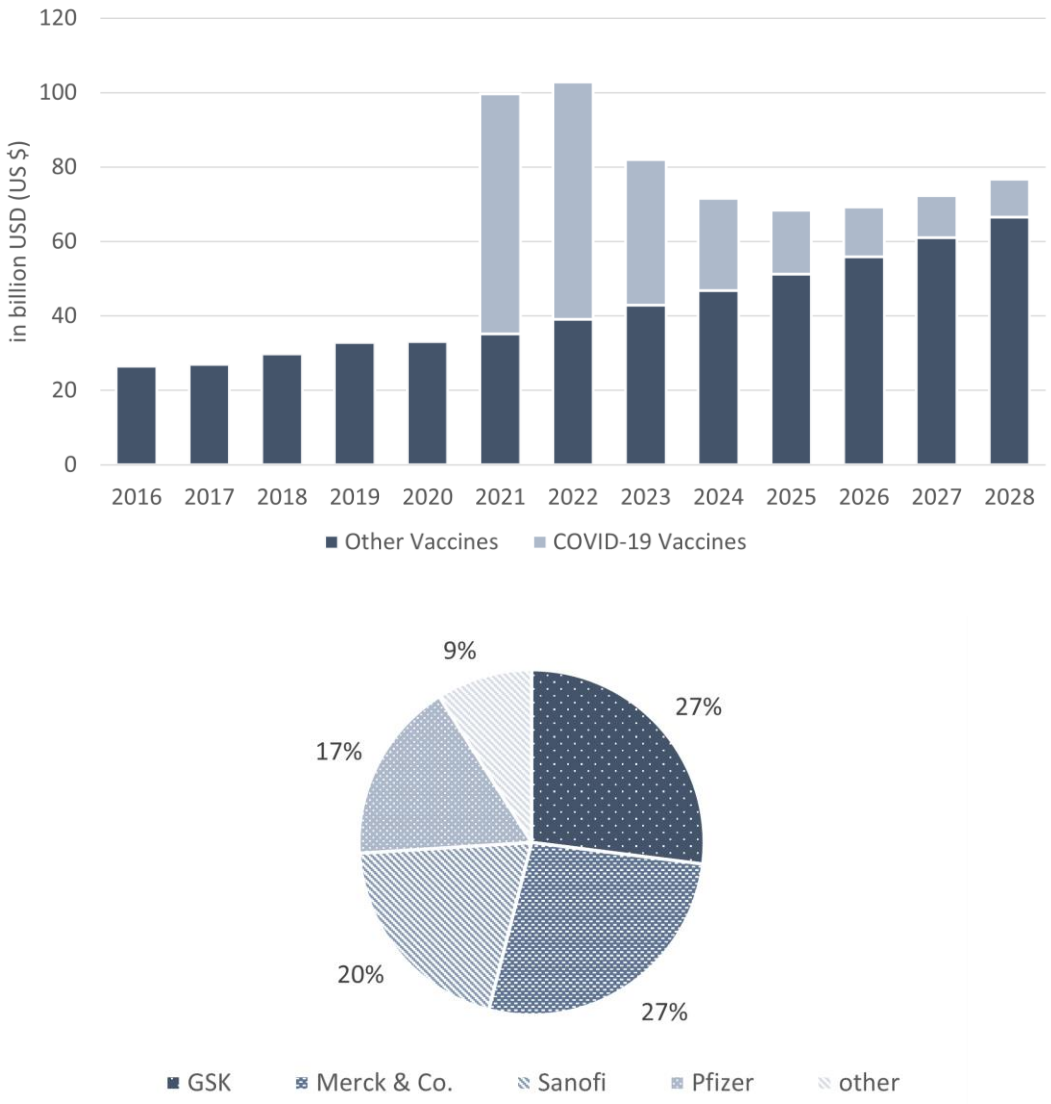


Figure 1. 1. Upper figure: The global revenues of vaccines, given in US dollars, from 2016 to the prospective year 2028. COVID-19 vaccines are separately indicated with a light blue color. Lower figure: Global market share, in percentages, between the leader companies, excluding COVID-19 vaccines (December 2023). Data source: Statista [1].

Infectious diseases are caused by various pathogens, including bacteria, viruses, parasites, or fungus. These pathogens consist of several distinct components, also known as antigens. When the human body is exposed to an antigen, the immune system responds by producing antibodies that specifically target the antigen and consequently eliminate the pathogen (see Figure 1.2A). Upon re-infection with the same pathogen, the immune system recognizes the antigen, resulting in a faster response and preventing the individual from severe sickness [8]. Vaccines evoke an immune response by presenting a foreign antigen to the immune system. As vaccines only contains a part of a pathogen or an inactivated/attenuated form, they induce an immune response without causing illness (see Figure 1.2B). In this way, the immune system

can respond faster once exposed to the real pathogen [8]. Different types of vaccines exist, such as whole pathogen (inactivated or attenuated), antigenic components (subunit) of pathogen, or nucleic acid vaccines [9].

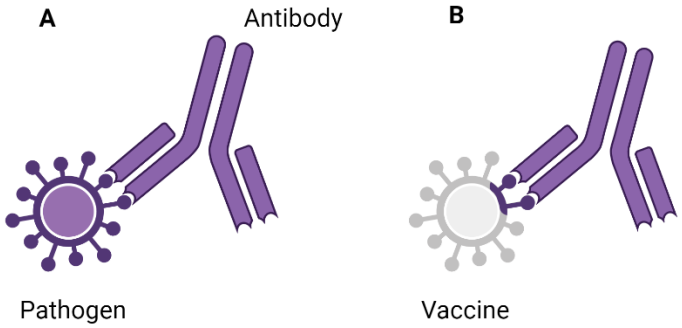


Figure 1. 2. A: Pathogen with specific antigens targeted by an antibody. B: The vaccine only consists of the antigen or a weakened version of the pathogen. Created with BioRender.com.

Developing a vaccine is a complex process that involves multiple clinical trials to establish the safety, efficacy, potency, and manufacturing consistency of the product [9]. Therefore, the production process of vaccines is crucial in determining the final product. A general vaccine production process consists of an upstream part, involving amplification of the antigen by fermentation or cell culture, and a downstream part, including the purification, and formulation of the product, see Figure 1.3 [10]. In the purification of vaccines, achieving very low contaminant concentrations is crucial to prevent issues such as high reactogenicity and unwanted immune responses. Downstream processing removes the majority of host cell impurities and process additives, aiming to achieve high product purity and yield [9]. Chromatography plays an essential role to achieve these very high product purities. However, the downstream process, particularly chromatography, represents a significant part of the total manufacturing costs [11]. While monoclonal antibodies (mAbs) share relatively similar properties, proteins subunit vaccines, for example, exhibit significant variation in physico-chemical properties, posing a greater challenge in standardizing the purification process [12]. As vaccines are administered to healthy people, the safety requirements are extremely high. This causes additional complexity for the process development, leading to a time-consuming and costly vaccine development [9, 13]. Hence, minimizing time-to-market is essential for the biopharmaceutical industry, yielding both life-saving outcomes and financial benefits. This emphasizes the importance of systematic, general, and efficient process development aiming to increase the process understanding and process control, and reduce process development times [14-16].

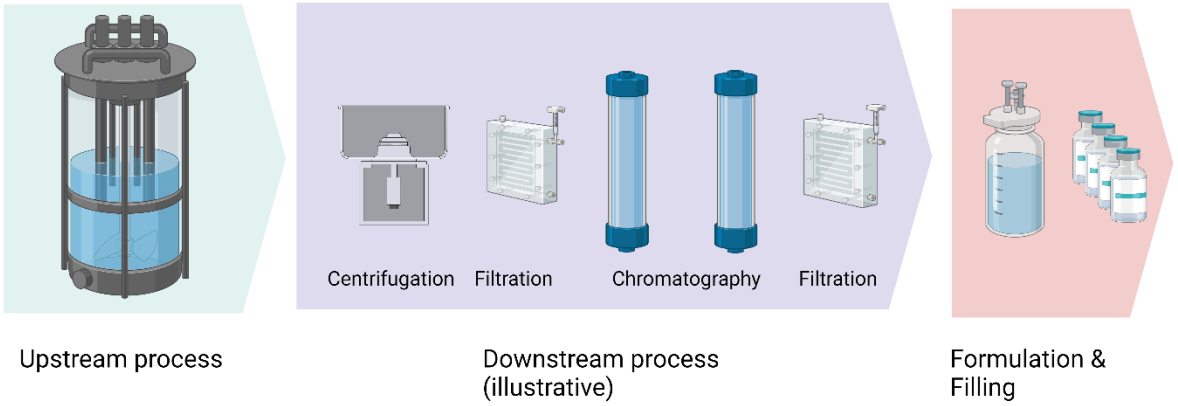


Figure 1. 3. General overview of a vaccine production process, starting with the upstream part consisting of a fermentation process. Followed by the downstream processing in which multiple separation techniques are used to purify the product. The final part consists of formulation of the product and filling it into small containers. Created with Biorender.com.

In recent and upcoming years, the biopharmaceutical industry is shifting towards more model-based process development, aligning with the Industry 4.0 for digitalization of the entire production process [17, 18]. Models are mathematical representations of real systems, allowing to run virtual experiments to enhance process and/or product understanding [19, 20]. Their applicability is versatile, models can function as digital twins for process control and monitoring purposes, or they can perform simulations for design or optimization purposes. Consequently, the use of modeling techniques reduces the need for extensive experimental effort and minimizes material costs.

Mechanistic models attempt to describe the physical phenomena occurring in a process or system [21, 22]. These models are based upon process knowledge and described in a mathematical form, including material and/or energy balances, as well as transport and thermodynamic phenomena. In order to describe the process, specific process related model parameters are needed, which can be determined experimentally or by physical correlations. An ongoing challenge for industry to adopt mechanistic modeling in their chromatography process development is the experimental determination of adsorption isotherms. A computational alternative is quantitative structure property relationship (QSPR), which aims to predict the retention behavior of proteins, or even adsorption isotherms parameters, based on the protein structure [23, 24]. Once the mechanistic model is developed, it needs to be validated, which involves the comparison between the modeled data and the experimental data to assess the model's accuracy in describing the real process. Subsequently, the validated mechanistic model can replace real experiments, for example, to screen operating conditions. However, for the final process design, an experimental verification is always required.

When mechanistic models for each step are present, an overall downstream process can be described. Eventually, the combination of purification steps will determine the overall process performance. Developing an entire downstream process involves numerous factors, including type and sequential order of the purification techniques, operating conditions, and costs [25, 26]. Especially in the early stages of development, it is desirable to identify the optimal downstream process considering costs, quality, and development time. The aim of optimization is to achieve specific objectives, such as obtaining a high yield (retaining the majority of product material), high purity (minimizing impurities in the product), and achieving a high productivity (producing a specified quantity within a certain time period, thereby reducing costs). The way a process can be optimized is by tuning the operating or design parameters, such as the salt concentration in the buffers, the duration of a purification step, and the size of the chromatography column, among others.

Optimizing the entire purification sequence at once by screening the overall design space is crucial for finding the optimal purification process [27]. This is because the most optimal purification process may not necessarily involve each unit operation performing at its individual optimum. Flowsheet optimization involves assessing all potential options, including the number, type and order of purification steps and their operating conditions, to purify the product [28]. Initially, a superstructure is designed, encompassing all possible process configurations, which are also referred as flowsheets. Subsequently, each flowsheet is optimized. Mechanistic models are utilized in the optimization, from the simulated chromatogram the process performances can be extracted, such as the yield, purity and productivity. Based on these outcomes, the optimization solver determines the necessary adjustments to the operating or design parameters, aiming to attain higher levels of yield, purity, and productivity. However, these mechanistic models can be speed-limiting, and this can be a significant disadvantage, particularly for flowsheet optimization purposes. Artificial Neural Networks (ANNs) can be used instead, serving as surrogate model of the mechanistic model and allowing for faster computations [29, 30]. The ANN functions as a ‘black-box’ model that is trained with certain input and output parameters, so only for the used in-and output parameters the ANN can be used [20]. The data needed to train, validate, and test the ANNs is obtained by running numerous simulations with the mechanistic model.

1.2. Project setting

This project ‘Computational modeling and optimization of biopharmaceutical downstream processes’ was funded by GlaxoSmithKline Biologicals S.A. (Belgium) and part of a collaboration between GlaxoSmithKline Biologicals S.A. (Belgium) and Technical University of Delft (The Netherlands). The collaboration aims to implement model-based high throughput

process development techniques into the end-to-end workflow of GSK's vaccine process development. This collaboration project focuses on *Escherichia coli* based recombinant vaccines and can be extended to other vaccines. Ultimately, this approach will contribute to reducing process development times by minimizing the experimental effort and enhancing process understanding through the application of mechanistic modeling. This collaboration comprises three PhD-projects. One of the projects focuses on developing experimental methods to characterize the host cell proteome and determine modeling parameters for this complex mixture. The other project focuses on QSPR modeling, in which protein structures are used to calculate physiochemical properties that correlate to specific retention behavior.

This PhD-project aimed to computationally describe and optimize the entire downstream process. Although, the project focus is to apply these methods to protein subunit vaccines, the developed methods can also be applied to other (bio)pharmaceuticals. Initially, mechanistic models were developed for ion-exchange chromatography, hydrophobic interaction chromatography and ultrafiltration/ diafiltration. Subsequently, an optimization software was built to perform flowsheet optimization, which supports decision-making in identifying the optimal purification process. Additionally, the use of ANNs to accelerate flowsheet optimization and various optimization strategies were explored. Finally, a multiscale modeling approach integrated QSPR and chromatographic mechanistic modeling, enabling the optimization of a cation exchange capture step based solely on knowledge of the protein structure.

1.3. Thesis outline

The *in silico* techniques developed and applied to downstream process case studies are described in this PhD thesis of which an overview is provided in Figure 1.4.

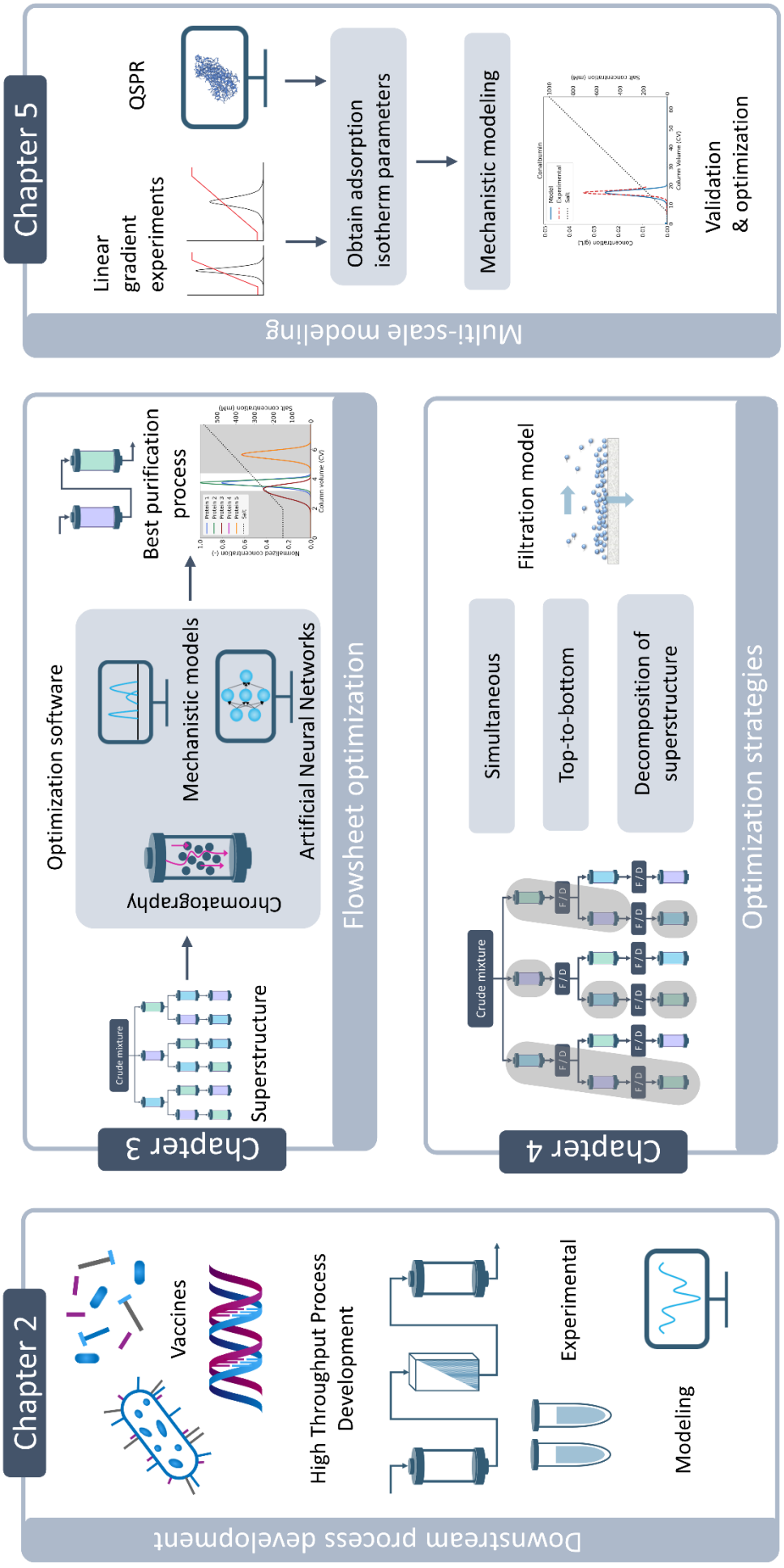


Figure 1. 4. Schematic overview of this dissertation.

In **Chapter 2**, current and future approaches to downstream process development in the biopharmaceutical industry are provided, with a particular focus on chromatography. The chapter discusses experimental-driven methods, such as, Design-of-Experiments and High Throughput Screening, followed by expert-knowledge approaches, including utilization of platform processes. The section of model-based downstream process development describes the different types of models, such as data-driven, mechanistic, and hybrid models. Subsequently, the chapter delves into research examples of high throughput process development studies. In the final section, with a focus on the future, the integration of Artificial Intelligence in process development is explored.

Advanced modeling techniques are employed to identify the global optimum within the overall design space. An optimization software is developed for performing complex flowsheet optimizations, utilizing mechanistic models to determine the process performances under specific conditions. In **Chapter 3**, ANNs are considered as an alternative to chromatographic mechanistic models during global flowsheet optimization, aiming to decrease the computational time and identify the best process sequence(s). A comparison of both modeling techniques during flowsheet optimization is conducted using a biopharmaceutical case study involving three types of chromatography with a maximum of three unit operations.

In chapter 3 a total of 15 flowsheets are evaluated, however, when considering more unit operations and different types, the number of flowsheets to be assessed increases. Hence, the optimization strategy becomes crucial in terms of time-efficiency, complexity, and outcome. In **Chapter 4**, we compare three optimization strategies, namely simultaneous, top-to-bottom, and superstructure decomposition to determine the most effective approach for complex flowsheet optimization. In this flowsheet optimization, we include an optional diafiltration mode between the chromatography steps, resulting in a total combination of 39 flowsheets to be assessed. Additionally, each optimization strategy is performed both by using mechanistic modeling and ANNs during the global optimization to also assess the difference in performance and duration between mechanistic models and ANNs. All strategies are successfully implemented and able to identify multiple optimal flowsheets. In summary, this chapter highlights the importance of various optimization strategies and modeling techniques for flowsheet optimizations.

As stated previously, this project aims to computationally describe the entire downstream process. However, a remaining bottleneck for mechanistic model implementations in industry is the experimental determination of adsorption isotherm parameters, which are needed as input parameters. In **Chapter 5**, we demonstrate a multiscale modeling approach in which we combine QSPR and mechanistic modeling techniques to optimize a cation exchange capture

step for an unseen protein. QSPR aims to correlate physicochemical properties with specific behavior, such as retention times or specific adsorption isotherm parameters. Once the database and QSPR model are developed, only the protein structure is needed to determine these specific model parameters and simulate the chromatographic process using mechanistic modeling. This multiscale modeling approach emphasizes the value of integrating diverse modeling techniques and, furthermore, reducing the dependence on wet-lab experiments, particularly in early phase process development.

Chapter 6 presents a final conclusion of this work and the key findings, together with prospects for future research.

1.4. References

[1] P.S. Azzopardi, et al., The unfinished agenda of communicable diseases among children and adolescents before the COVID-19 pandemic, 1990–2019: a systematic analysis of the Global Burden of Disease Study 2019, The Lancet 402(10398) (2023) 313-335. https://doi.org/https://doi.org/10.1016/S0140-6736(23)00860-7.

[2] G. Jagschies, E. Lindskog, K. Łacki, P. Galliher, Biopharmaceutical Processing: Development, Design, and Implementation of Manufacturing Processes, 2018.

[3] F.E. Andre, R. Booy, H.L. Bock, J. Clemens, S.K. Datta, T.J. John, B.W. Lee, S. Lolekha, H. Peltola, T.A. Ruff, M. Santosham, H.J. Schmitt, Vaccination greatly reduces disease, disability, death and inequity worldwide, Bulletin of the World Health Organization 86(2) (2008) 140-146. https://doi.org/10.2471/BLT.07.040089.

[4] C.M.C. Rodrigues, S.A. Plotkin, Impact of Vaccines; Health, Economic and Social Perspectives, Frontiers in Microbiology 11(1526) (2020). https://doi.org/10.3389/fmicb.2020.01526.

[5] Statista, Vaccines - Worldwide. (n.d.). https://www.statista.com/outlook/hmo/pharmaceuticals/vaccines/worldwide?currency=USD&locale=en

[6] G. Walsh, E. Walsh, Biopharmaceutical benchmarks 2022, Nat Biotechnol 40(12) (2022) 1722-1760. https://doi.org/10.1038/s41587-022-01582-x.

[7] UNICEF, Supply division.

[8] W.H. Organization, How do vaccines work?, 2020. https://www.who.int/news-room/feature-stories/detail/how-do-vaccines-work.

[9] E.P. Wen, R. Ellis, N.S. Pujar, Vaccine Development and Manufacturing, Wiley2014.

[10] S.B. Carvalho, C. Peixoto, M.J.T. Carrondo, R.J.S. Silva, Downstream processing for influenza vaccines and candidates: An update, Biotechnology and Bioengineering n/a(n/a) (2021). https://doi.org/https://doi.org/10.1002/bit.27803.

[11] K.M. Łącki, Chapter 16 - Introduction to Preparative Protein Chromatography, in: G. Jagschies, E. Lindskog, K. Łącki, P. Galliher (Eds.), Biopharmaceutical Processing, Elsevier2018, pp. 319-366. https://doi.org/https://doi.org/10.1016/B978-0-08-100623-8.00016-5.

[12] Y.-p. Yang, T. D'Amore, Protein Subunit Vaccine Purification, Vaccine Development and Manufacturing2014, pp. 181-216. https://doi.org/https://doi.org/10.1002/9781118870914.ch6.

[13] M. Zhao, M. Vandersluis, J. Stout, U. Haupts, M. Sanders, R. Jacquemart, Affinity chromatography for vaccines manufacturing: Finally ready for prime time?, Vaccine 37(36) (2019) 5491-5503. https://doi.org/https://doi.org/10.1016/j.vaccine.2018.02.090.

[14] L.X. Yu, Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control, Pharmaceutical Research 25(4) (2008) 781-791. https://doi.org/https://doi.org/10.1007/s11095-007-9511-1.

[15] A.S. Rathore, Quality by Design (QbD)-Based Process Development for Purification of a Biotherapeutic, Trends Biotechnol 34(5) (2016) 358-370. https://doi.org/https://doi.org/10.1016/j.tibtech.2016.01.003.

[16] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: knowledge-based chromatographic process development, Trends Biotechnol 32(4) (2014) 210-220. https://doi.org/https://doi.org/10.1016/j.tibtech.2014.02.001.

[17] M. Bisschops, L. Cameron, Process Intensification and Industry 4.0: Mutually Enabling Trends, Process Control, Intensification, and Digitalisation in Continuous Biomanufacturing2022, pp. 209-229. https://doi.org/https://doi.org/10.1002/9783527827343.ch7.

[18] I.C. Reinhardt, D.J.C. Oliveira, D.D.T. Ring, Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector, Journal of Industrial Information Integration 18 (2020) 100131. https://doi.org/https://doi.org/10.1016/j.jii.2020.100131.

[19] P.-A. Muller, F. Fondement, B. Baudry, B. Combemale, Modeling modeling modeling, Software & Systems Modeling 11(3) (2012) 347-359. https://doi.org/10.1007/s10270-010-0172-x.

[20] M. von Stosch, R. Oliveira, J. Peres, S.F. de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput Chem Eng 60 (2014) 86-101. https://doi.org/https://doi.org/10.1016/j.compchemeng.2013.08.008.

[21] D. Solle, B. Hitzmann, C. Herwig, M. Pereira Remelhe, S. Ulonska, L. Wuerth, A. Prata, T. Steckenreiter, Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation, Chemie Ingenieur Technik 89(5) (2017) 542-561. https://doi.org/https://doi.org/10.1002/cite.201600175.

[22] A. Felinger, G. Guiochon, Comparison of the Kinetic Models of Linear Chromatography, Chromatographia 60(1) (2004) S175-S180. https://doi.org/https://doi.org/10.1365/s10337-004-0288-7.

[23] D. Saleh, R. Hess, M. Ahlers-Hesse, F. Rischawy, G. Wang, J.-H. Grosch, T. Schwab, S. Kluters, J. Studts, J. Hubbuch, A multiscale modeling method for therapeutic antibodies in ion exchange chromatography, Biotechnology and Bioengineering 120(1) (2023) 125-138. https://doi.org/https://doi.org/10.1002/bit.28258.

[24] C.B. Mazza, N. Sukumar, C.M. Breneman, S.M. Cramer, Prediction of Protein Retention in Ion-Exchange Systems Using Molecular Descriptors Obtained from Crystal Structure,

Analytical Chemistry 73(22) (2001) 5457-5461. https://doi.org/https://doi.org/10.1021/ac010797s.

[25] T.C. Huuk, T. Hahn, A. Osberghaus, J. Hubbuch, Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography, Sep Purif Technol 136 (2014) 207-222. https://doi.org/https://doi.org/10.1016/j.seppur.2014.09.012.

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

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

[28] H. Yeomans, I.E. Grossmann, A systematic modeling framework of superstructure optimization in process synthesis, Comput Chem Eng 23(6) (1999) 709-731. https://doi.org/https://doi.org/10.1016/S0098-1354(99)00003-4.

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

[30] D. Nagrath, A. Messac, W.B. B, M.C. S, A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes, Biotechnology Progress 20(1) (2004) 162-178. https://doi.org/https://doi.org/10.1021/bp034026g.

Chapter 2

Recent advances to accelerate purification process development: a review with a focus on vaccines

The safety requirements for vaccines are extremely high since they are administered to healthy people. For that reason, vaccine development is time-consuming and very expensive. Reducing time-to-market is key for pharmaceutical companies, saving lives and money. Therefore the need is raised for systematic, general and efficient process development strategies to shorten development times and enhance process understanding. High throughput technologies tremendously increased the volume of process-related data available and, combined with statistical and mechanistic modeling, new high throughput process development (HTPD) approaches evolved. The introduction of model-based HTPD enabled faster and broader screening of conditions, and furthermore increased knowledge. Model-based HTPD has particularly been important for chromatography, which is a crucial separation technique to attain high purities. This review provides an overview of downstream process development strategies and tools used within the (bio)pharmaceutical industry, focusing attention on (protein subunit) vaccine purification processes. Subsequent high throughput process development and other combinatorial approaches are discussed and compared according to their experimental effort and understanding. Within a growing sea of information, novel modeling tools and artificial intelligence (AI) gain importance for finding patterns behind the data and thereby acquiring a deeper process understanding.

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

2.1. Introduction

The COVID-19 pandemic has engulfed the world, which has already cost over millions of lives and is still infecting hundreds of thousands of people every day, one and a half year after the first outbreak in December 2019. More than ever the world is aware of the value of vaccination, contributing to improved public health, reduced healthcare costs, economic growth, travel safety and prolonged life expectancy [1, 2]. In general, vaccination is estimated to prevent 2-3 million childhood and almost 6 million adult deaths annually [1, 3]. Recently, the WHO published an action plan making vaccination available to everyone in the world and promoting innovation within the vaccine industry [4].

The downstream process plays a key role in reducing contaminant concentrations in vaccines to very low values. This prevents for example high reactogenicity and unwanted immune responses, and guarantees the safety and efficacy of the vaccine. Designing a vaccine purification process is accompanied with many decisions, such as type and sequential order of purification techniques, conditions, costs, and other performance measurements [5]. Additionally, optimization of a single unit operation and overall purification sequence is important, whereas small variations of conditions in one step may affect the subsequent unit operation performance. High safety and purity demands lead to increased complexity of the vaccine purification process. This, often along with a low productivity and process capability, makes the downstream process very expensive in both costs and time [6, 7]. One of the main challenges in developing vaccine purification processes is the separation of critical impurities closely related to the product, such as host cell proteins (HCPs) to a protein-antigen vaccine or genomic DNA or RNA to a DNA or RNA-based vaccine, respectively. Another challenge is the preservation of the antigen structure during the purification process, as well as the antigen stability, as most antigens are vulnerable to temperature, pH or salt concentration changes.

Fast vaccine process development is of utmost importance in light of infectious outbreaks and pushing competitive market, which highly depends on its design strategy for the purification process [6]. Traditionally, vaccines are developed within 10 – 15 years, hence pharmaceutical companies desire to reduce the process development time drastically in every aspect. One of the reasons the first SARS-CoV-2 vaccines could be developed within 1-2 years, is the employment of an accelerated development timeline due to parallelization of phases instead of sequential development [8]. Additional reasons for such a quick development are the application of previous knowledge and production processes from related viruses and existing vaccines (i.e. platform knowledge), and widely available government funding enabling parallelization, risk-taking and fast regulatory reviewing [9].

The 'quality by design' (QbD) paradigm [10, 11] made the pharmaceutical industry shift from a trial-and-error approach towards a more comprehensive, systematic, and efficient approach, with the purpose to increase process understanding and process control [12-16]. The implementation of high-throughput process development (HTPD) approaches contributes to faster and more efficient process developments, additionally decreasing material consumption and improving cost-effectiveness [16]. HTPD is a combinatorial approach of both high throughput experimentation (HTE) and modeling techniques. Recently Sao Pedro et al. outlined the areas of major problems (e.g. cell culture, filtration and analytical tools) within HTPD, along with suggested solutions (microfluidics, modeling and Process Analytical Technologies (PAT)) for the purpose of integrated and continuous bio manufacturing [17]. Although this review is not focused on continuous biomanufacturing, the current limitations of HTPD are likewise applicable to the vaccine purification process development.

Vaccine purification processes can differ enormously from each other as they depend strongly on the type of vaccine and crude starting material/host organism (e.g. fertilized eggs, bacterial-, mammalian-, and insect cells). Carvalho et al. pointed out the influence of vaccine types on downstream process strategies and described into detail each vaccine purification step with a focus on influenza vaccines [18]. A general overview of vaccine types is shown in Figure 2.1, being classified either as whole pathogen (inactivated or attenuated), antigenic components (subunit) of pathogen or nucleic acid vaccines, though slightly different classifications have also been reported.

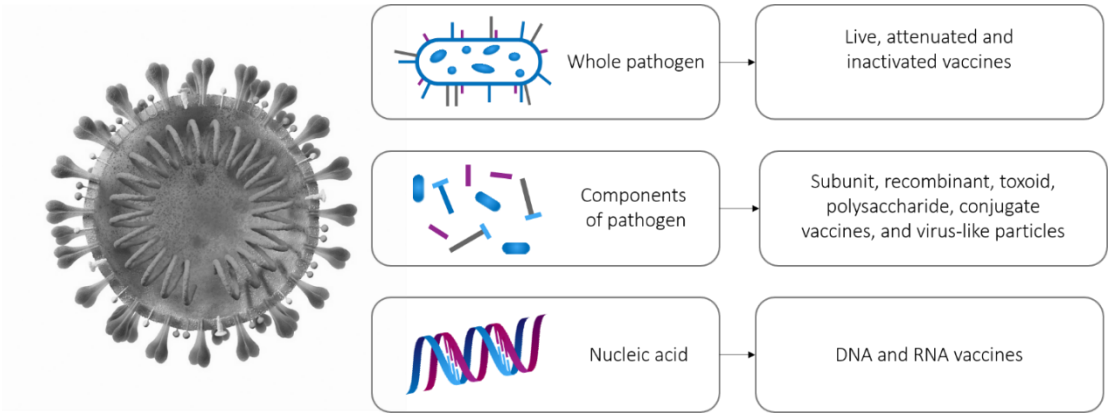


Figure 2. 1. Types of vaccines classified in whole pathogen, antigenic components of pathogen and nucleic acid vaccines [2-5].

In order to preserve the genetic stability of live and inactivated vaccines, the downstream process consist of only a few steps. The purification of protein recombinant or subunit vaccines involves a complex purification challenge because of the presence of HCPs closely-

related to the target protein [6]. Recently Jones et al. pointed out the concerns of high-risk HCPs and recommended a strategy for monitoring and eliminating the known impurities [19]. Despite the great variance between different protein subunit vaccine downstream processes, the generic order of purification steps is similar as shown in Figure 2.2. If the antigen (product of interest) is produced intracellular the purification process requires a cell lysis step, while this step is not needed if the antigen is produced extracellular. Detailed purification schemes for certain vaccine types are outside the scope of this paper and can be found elsewhere. For example, Josefsberg and Buckland [20] described the production process of several virus-based conjugate and DNA vaccines, while Abdulrahman and Ghanem [21] summarized the most recent advances in the purification of plasmid DNA vaccines. In the book of Wen et al., viral vaccines purification and protein subunit vaccines purification are described into more detail [6, 22, 23].

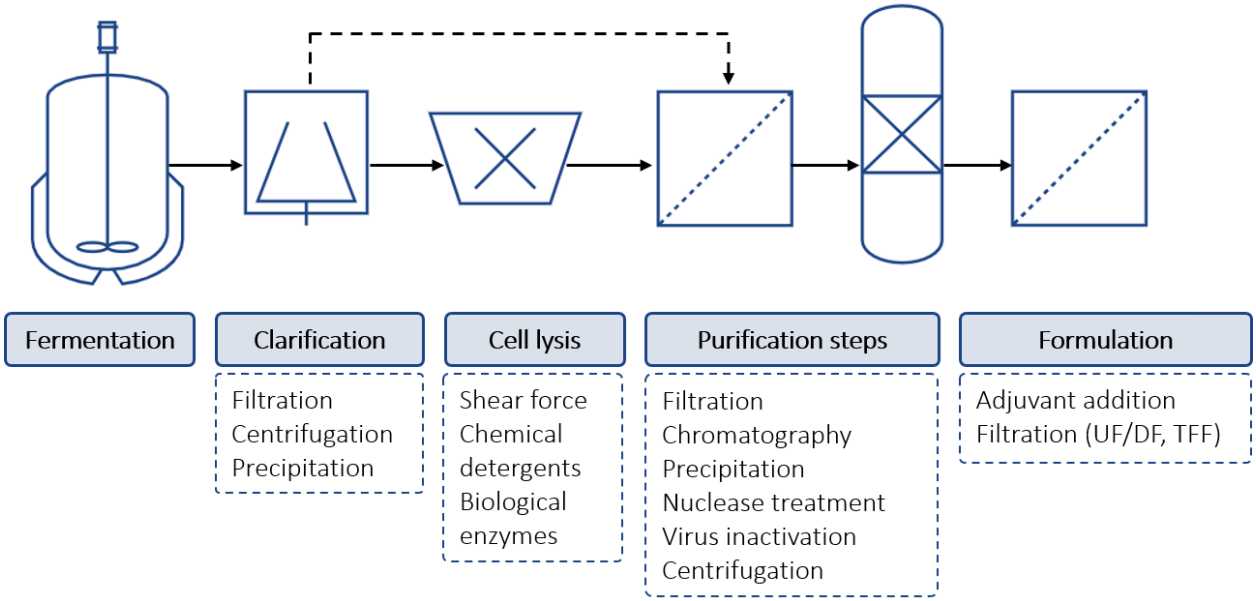


Figure 2. 2. General process flowsheet for vaccines including the upstream and downstream part, from fermentation to the last formulating steps. The optional processing techniques for different types of vaccines are given below each unit operation. The solid line represent a purification process in which the antigen is produced intracellular, including the cell lysis. The dashed line shows a purification process for extracellular products excluding cell lysis.

Most of the current vaccine development approaches are based on design of experiments (DoE), in which multiple factors are changed simultaneously to evaluate the underlying interactions, thereby obtaining a multidimensional model that correlates the effects of various factors on the critical quality attributes (CQA), which is an essential aspect within QbD guidelines [14, 24]. However, the existing vaccine process development strategy requires high experimental effort and little process understanding is gained through it. Moreover, the

sequential determination of purification steps and individual process optimizations might lead to a suboptimal process design with respect to the objective, such as yield or costs [25-27]. A standardized approach, also known as platform process, as established for monoclonal antibodies (mAbs) is yet missing, mainly due to the large diversity between vaccine types [28]. Even when considering only protein subunit vaccines, already a very diverse range of proteins can be found due to a variety of expression systems.

A platform process for specific vaccine types would be highly beneficial in terms of process development time, knowledge, resources, costs and regulatory aspect [7]. Another often complicated task is the precise quantitatively measurement and characterization of virus or bacterial particles, further complicated by the lack of rapid analytical technologies [7, 22]. A trend within the QbD initiative is the use of PAT, allowing real-time measurements to ensure consistent product quality and performance, besides providing a better understanding of the process [14]. Mechanistic models rely on physical processes occurring during a certain separation step and can therefore be of great merit to the process understanding, but also decrease experimental effort and allow to perform processes on different scales *in silico*. The use of AI techniques could eliminate shortcomings within the modeling area and bring modeling techniques to a higher level of applicability and usability.

This review presents modern and future downstream process development approaches and their application in (bio)pharmaceutical industry with a focus on chromatography, as this is the main purification technique for protein subunit vaccines. This paper aims to show the evolvement of model-based high throughput process development approaches through the use of more advanced modeling techniques, such as empirical, mechanistic and hybrid modeling. The applicability and benefit using these methods are supported by case studies from industry and academia.

2.2. Downstream process development methods

The overall goal of process development is to design the optimal purification process, by means of achieving purity targets at minimum costs and time efforts, while at the same time adhering to all regulatory requirements. Currently, vaccine development employs mostly DoE-based methods, though it could benefit from more advanced model-based process development approaches, which are already used in other biopharmaceutical branches, such as for the purification of mAbs. Figure 2.3 shows two types of process development approaches, the DoE-based method and a modeling-based method. In the following section,

process development approaches are described briefly. More comprehensive reviews on this topic can be found elsewhere [16, 29].

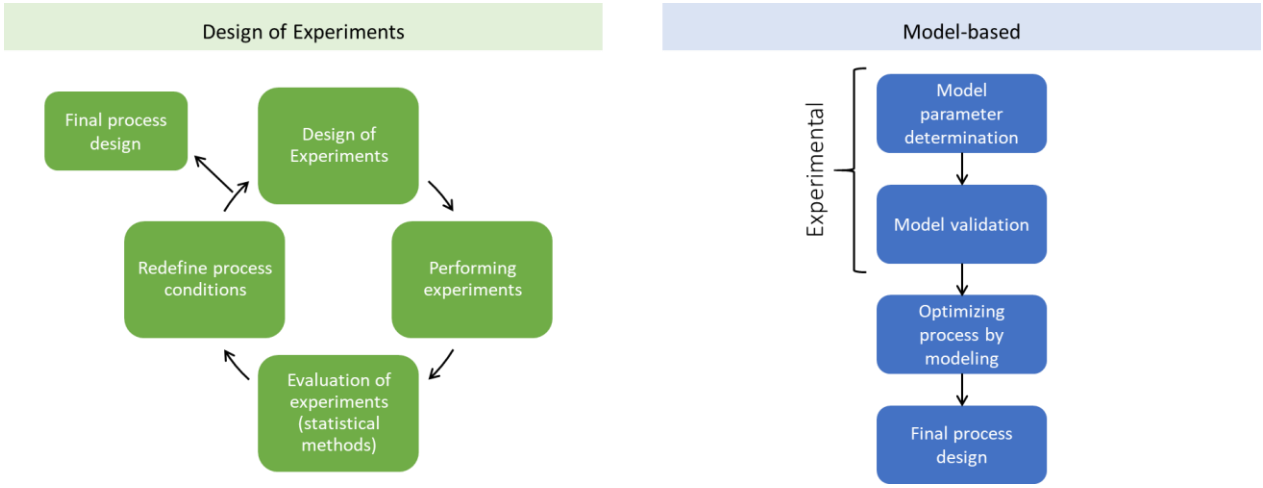


Figure 2. 3. Overview of two different process development approaches. Left: Design of Experiments (DoE) approach, which performs experiments based on statistical tools and evaluating the results by statistical analysis. This approach is commonly applied within biopharmaceutical industry. Right: Model-based process development approach, which employs targeted experiments to determine model input parameters such as isotherm parameters and column parameters. The model has to be validated before performing the optimization.

2.2.1. Experimental driven downstream process development

2.2.1.1. One-factor-at-a-time and Design of Experiments

One-factor-at-a-time (OFAT) is a more traditional approach in which one factor is changed during a series of experiments while the other factors are kept constant. In this method dependencies between factors are neglected and therefore discovery of the optimum is rather difficult and quite inefficient [30]. For that reason, the biopharmaceutical industry shifted more than a decade ago to the statistics-based DoE approach to design and analyze experiments, thereby obtaining more valuable information by conducting less experiments. The classical DoE-method is factorial design. Experiments are performed on all possible combinations of factors with the purpose to identify effects of each factor as well as interactions between factors on the response. An improvement on the classical DoE-screenings is definitive screening design which estimates the curvature effects and enables separation of factors having a significant impact on the response from the factors having negligible effects. Oher methods offering a three level multifactorial design are for example Box-Behnken [31] or central composite designs. Hibbert extensively described the most common used DoE methods with a focus on chromatography [32, 33]. Various DoE software

are available nowadays, such as Design-Expert, Modde and JMP, though other statistical software, like R, SPSS and various Python packages, can also be used for DoE purposes.

2.2.1.2. Parameter acquisition for modeling purposes

An alternative experimental strategy is to determine parameters that serve as input for mechanistic or physical models. The use of mechanistic models has been established decades ago and is nowadays widely adopted by chemical industry, where some processes are even designed entirely *in silico* [34]. Only recently, biopharmaceutical and vaccine industry initiated this strategy into their process development, in which the major challenge is often the complex feed mixture containing the product of interest (e.g. antigen) together with thousands of proteins and impurities [23]. This is probably why mechanistic modeling together with parameter acquisition has not been widely adopted yet, as it is nearly impossible to experimentally determine and model thousands of proteins and impurities. However, HTE made it worthwhile to determine model parameters even for more complex mixtures [35-37]. Noteworthy, a validated model increases process understanding and enables to optimize the process *in silico*, resulting in time, material and costs savings [38]. For chromatographic purposes, as this is the main purification technique in protein subunit vaccines, the adsorption isotherm parameters describing the binding behavior of components to the solid phase, are of utmost importance. Experimental determination of adsorption equilibria is required to establish the isotherm parameters and can be obtained by batch adsorption experiments [36, 39-42], frontal analysis, isocratic elution or linear gradient elution [41, 43, 44] or by making use of inverse techniques, which minimize the difference between experimental and simulated elution profiles by tuning certain parameters [36, 44, 45]. Besides isotherm determination, column and resin characteristics must also be obtained in order to acquire a validated model, however these are more straightforwardly obtained [41].

2.2.1.3. High Throughput Screening (HTS)

The introduction of liquid handling stations (LHS), about two decades ago, allowed the acceleration of conducting experiments, also known as HTE or HTS. Due to automation, miniaturization and parallelization it became viable to create large data sets while using a reduced amount of sample volume and resources within a shorter time-frame [46, 47]. Another benefit of automation is the lowered variability and superior precision [48]. Nowadays, LHS is a widely applied technique in both academia and industry and reduces the process development time significantly [49-51]. As LHS allows to screen more conditions, it is more feasible to find optimal conditions for a purification process. Apart from the system's benefits there are certainly also some disadvantages pointed in literature [52, 53]. For example, the LHS's limitation in accurately mimicking the flow distributions of process

columns [49]. HTS requires high understanding of efficient experimental design in order to make optimal use of the system, therefore it is rather a tool to be used than an approach on its own.

2.2.2. Expert-knowledge driven downstream process development

2.2.2.1. Universal

Rules of thumb, available knowledge and experience of previous processes are the basis for expert knowledge or heuristic approaches to design new production processes [29, 54]. Using expert insights is easy to apply and can speed up the process design by eliminating combinations of unit operations with less promising results [55]. Asenjo et al. developed an expert system focused on downstream protein processes; this software uses databases consisting of expert knowledge on universal process designs (heuristics) to support and accelerate decision-making for the selection of a sequence of unit operations [54, 56]. Several handbooks, like Sofer & Hagel [57] and GE healthcare [58], outline general design heuristics extensively. Most vaccine purification processes are also based upon heuristics, as for example the purification of hepatitis A virus from mammalian cell cultures, in which the first step involves a low-cost anion-exchange chromatography to capture the product and remove a substantial amount of impurities and the last step of the downstream process a polishing and desalting step using size-exclusion chromatography [22, 57, 59]. A general example that is almost entirely based on knowledge are platform processes as explained into more detail in the next paragraph.

2.2.2.2. Platform process

Platform processes are used as ‘templates’ for designing an entire purification sequence for a specific type of molecule, utilizing a pre-established series of unit operations [29]. The platform instructions provide details of the operating conditions for each unit operation, corresponding to the overall purification process. One of the key advantages is a reduced process development time, regulatory aspect and resources for similar molecules and accordingly decreased time-to-market and validation effort [57]. Moreover, the platform documents can be shared and aligned among not only different departments, but also across different manufacturing sites, serving as a site-independent process [60]. The platform process approach is most suited for biopharmaceuticals with similar characteristics and thus purification steps [28, 57]. For example, mAbs are relatively well defined and platform processes are used to establish similar purification processes for new mAbs variants. Detailed information about process-related contaminants such as persistent HCPs and other impurities for the corresponding cell culture, i.e. CHO and hybridism are known [60]. The order of purification steps includes protein A chromatography, low pH viral inactivation, IEX

chromatography polishing steps, viral filtration, and ultrafiltration/diafiltration. Only small changes are required in the purification process conditions to determine a new mAb variant purification process. Other potential candidates for platform approaches could be pDNA vaccines and influenza vaccines, both having similar properties and purification steps [21, 57]. However, mAbs are relatively similar to each other in their properties, while protein subunit vaccines vary greatly in their appearance, making it more difficult to standardize the purification process.

2.2.3. Model-based downstream process development

In process engineering models play an important role, they aim to represent a real system in an abstracted mathematical format [61, 62]. Bézivin and Gerbé defined a model as "a simplification of a system built with an intended goal in mind. The model should be able to answer questions in place of the actual system" [63]. The intended goal related to process engineering could be for example, control, simulation, design, monitoring or optimization. Depending on the goal, different models can be appropriate [64]. Models help to understand complex problems and could provide potential solutions if the model is an adequate representation of the modeled system's features of interest [65]. Running the model with a given set of parameters is a simulation and hence an inexpensive and safe way to run a virtual experiment [66]. For that reason, the number of experiments in laboratory can be reduced and/or designed more efficiently, thereby reducing time and material consumption. Although, using models sounds attractive and promising, it does cost time, effort and knowledge to develop decent models that are able to fulfill the desired purposes. Moreover, there is a lack of educated people in this area that can develop and maintain scientific-, and engineering software. Within the near future, it is expected that more process engineers or scientist are familiar with modeling, because most technical related studies provide programming and data-processing courses nowadays. In order to build a model two main resources are essential, knowledge of the process, translated into laws of nature, and the collection of data obtained from the real system [66]. In process engineering, a distinction can be made between first-principles, mechanistic or knowledge-driven models and data-driven or empirical models, respectively known as transparent white-box and less transparent black-box models [61]. A combination of both is named hybrid semi-parametric models. An overview of the main advantages and disadvantages is given in Table 2.1.

Table 2. 1. Overview of the main advantages and disadvantageous of different modeling approaches.

	Advantages	Shortcomings
Data-driven models	- Requires no or little process understanding in advance - Takes less effort/time to develop the model - Easy to use and understand	- Only valid in a predefined measured region - Extrapolation generally not applicable - Parameters have often no physical meaning - Data-collection might be an issue for the application and generalization of data-driven models in biomanufacturing industry
Mechanistic models	- Allows extrapolation and exploration of conditions beyond measured results - Acquires process understanding - Parameters have a physical meaning	- Requires process understanding in advance - Complex to develop and hence time and effort - Determination of model parameters can be difficult
Hybrid models	- Eliminate drawbacks of certain modeling approaches - Improved model accuracy and extrapolation properties - Less data is required compared to purely data-driven models	- Requires additional effort, time and knowledge to develop hybrid models - Data-collection can be challenging

2.2.3.1. Data-driven models

Data-driven or empirical models attempt to describe the input-output relation based upon observed experiments within a predefined design space, such as artificial neural networks (ANN), statistical and regression models [64]. The biopharmaceutical industry often makes use of statistical models, either by executing a predefined set of experiments using DoE and an appropriate statistical data analysis method, such as response surface methodology (RSM), or by employing a multivariate data analysis using an existing dataset [67]. RSM is a well-known empirical model and describes the relation of a response between different tested factors within a DoE, and produces a model describing the mathematical relationship [32]. This statistical (black-box) model solely observes the factor-to-response correlation without gaining fundamental mechanistic (physiochemical) understanding of the estimated parameters. By making use of DoE and regression analysis through first- and second order polynomials the optimum input combination can be estimated [68]. However, fitting data to second order polynomials is a major drawback of RSM, as frequently not all curvatures within the systems can be described by the second order polynomial [69]. DoE in combination with empirical modeling has been widely applied to design downstream purification processes in biopharmaceutical industry and academia [70-72]. The effect of high-salt solution on RNA precipitation and pDNA recovery was investigated using DoE and linear regression models [71]. And more recently, Chiang et al. evaluated the impact of chromatographic parameters on virus clearance when switching from a single to multicolumn operation utilizing DoE [73]. A major limitation of data-driven models is that they are merely valid in a defined region of measured variables and only able to predict variables within that region, making extrapolation generally highly inaccurate. Moreover, little process knowledge can be extracted, because the parameters are often just correlations [74]. On the other hand, data-driven modeling requires no process understanding in advance and is less time consuming compared to mechanistic modeling [74].

2.2.3.2. Mechanistic models

Mechanistic, first-principle, or knowledge-driven models attempt to describe the inner mechanisms and phenomena occurring in a process or system based upon knowledge about the process. These models consist of material and/or energy balances together with transport and thermodynamic phenomena and have a fixed structure, meaning the parameters might have a physical interpretation [74]. The model parameters are estimated by experimental data or physical correlations. The physical processes occurring during a purification process can be translated into mathematical simulation models. A validated mechanistic model allows to explore various conditions *in silico* and therefore enables to acquire optimum operating conditions efficiently [75]. The phenomena taking place inside a chromatographic column are

well described in literature, Ruthven extensively outlined the dynamics and adsorption processes [76]. Kinetic or rate models are most common in practice, including dispersive factors, like mass transfer and dispersion effects, and equilibrium factors, such as adsorption isotherms, ionic dissociation and intermolecular association [41]. The three most prominent kinetic models are lumped kinetic model, lumped pore model, and general rate model, which are listed in order of complexity. The main difference between these models is the degree of covering pore diffusion effects [77]. However, it applies for all mechanistic models that isotherm parameters are crucial as explained previously in section 2.1.2, for which numerous binding models exist, such as linear, Langmuir [78], steric mass action [43] and mixed-mode [39]. The utilization of chromatographic models varies from process synthesis, optimization and control [79-85] to scale-up, resin selection and robustness checks [86-89]. One step further is the simulation of a combination of integrated chromatography and other conditioning steps to find the overall optimum purification process [5, 25, 90-93]. Nowadays, various commercial software of chromatographic mechanistic models are available, e.g. GoSilico (now part of Cytivia, and formally known as ChromX) [94], Aspen Chromatography, DelftChrom, CADET [95] and ChromaTech [96].

Alternative *in silico* methods for adsorption experiments have been investigated for several years. Molecular dynamics simulations attempt to describe the interaction between resin-proteins on a detailed atomic level [97-99]. Quantitative structure activity relationships (QSAR) combine molecular properties with empirical modeling to find correlations amongst retention behavior and protein surface properties [100-102]. This kind of molecular modeling can be used to predict the retention behavior of proteins on resins to reduce process development times [103]. However, often detailed information is required about each component, such as amino acid sequence or crystal structure and also a large amount of experiments [75].

Mechanistic models can explore conditions over a wide range and even beyond the observed measured results, possessing an increased extrapolation capability compared to data-driven models [74]. This contributes to process understanding, which is line with the QbD initiative, although mechanistic modeling also requires physical understanding. The major drawback of knowledge-based models is their complexity, hence requiring more development time compared to data-driven models.

2.2.3.3. Hybrid (semi-parametric) models

Hybrid (semi-parametric) modeling combines parametric (i.e. first principle-, mechanistic-, and knowledge-based models) with nonparametric (i.e. data-driven models) in order to eliminate drawbacks of individual approaches and get the best out of both [61]. Von Stosch et

al. extensively reviewed the hybrid semi-parametric modeling framework and the various applications in (bio)chemical engineering concerning process monitoring, control, optimization, model-reduction and scale-up. The parametric and nonparametric models can be configured in series or parallel, depending on the scope of the model. Usually a parallel mode is recommended when the mechanistic (white-box) model performance is limited or insufficiently accurate and the addition of a nonparametric (black-box) model may improve the estimations, Figure 2.4C. A serial approach is often utilized for reducing complexity of mechanistic models by determining parameters using nonparametric models, Figure 2.4A, or when the results of mechanistic models function as an input for nonparametric models, Figure 2.4B [61].

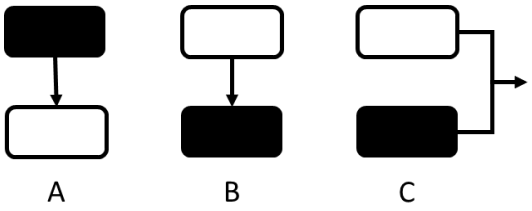


Figure 2. 4. Hybrid modeling configurations, white-boxes represent mechanistic/first-principle models and black-boxes represent data-driven models [2]. Serial approach (A, B) and parallel mode (C).

The usefulness of hybrid modeling lies within its capability to cost-effectively and efficiently solve a complex problem and develop a model. Other advantages are an improved model accuracy, transparency and extrapolation properties, besides gaining a broader process understanding [74]. However, the challenge is knowing in what manner different type of models can be combined to develop a hybrid model. Therefore, thorough understanding on both data-driven and mechanistic models is desired, as well as knowledge to acquire the correct data. Hybrid modeling is gaining more interest in both industry and academia, and seems to be a promising approach to overcome deficiencies in data-driven-, and mechanistic models.

2.3. High Throughput Process Development

2.3.1. Single or double purification steps

Hybrid process development approaches combine experimental and modeling tools to design a process. After the introduction of the LHS, hybrid approaches gained a special interest as LHS enabled experimentation in high throughput manner. Utilizing HTE in relation to process design is known as HTPD, combining HTS with empirical or mechanistic modeling is named

model-based HTPD [16, 38]. The implementation of HTPD pursues the QbD paradigm in terms of process and product understanding, hence contributing to high and stable product quality as well as process robustness [47]. The establishment of HTPD arose about 15 years ago [51, 55, 104, 105] and evolved ever since as an efficient and cost-effective method broadly acknowledgement by industry [15]. (Model-based) HTPD can be applied in various development stages and for different purposes, like resin and solubility screenings, design-space definition, risk-assessment, process robustness and control. In the review of Baumann and Hubbuch, several commercial miniaturized HT-suitable systems in both up and downstream process development are described [29]. The technical review of Lacki outlined the most frequently used chromatography HT equipment, such as microtiter filter plates, prefilled pipette tips and robocolumns, nowadays ranging from 50-600 µL [52]. Here, HTPD research from academic and industrial researchers are discussed. One can find more details on these and other HTPD approaches in Table 2.2. Depending on the purpose of the research a different HTDP approach is suitable, for example resin selection usually goes together with the use of empirical models while mechanistic modeling is preferred for an overall process design including multiple purification steps.

Bhambure and Rathore proved a tremendous increase in productivity (170x higher) utilizing a HTPD platform (2 and 6 µL resin volume) against the traditional laboratory scale (0.5 mL resin volume) for defining the characterization space of an ion exchange chromatography step using DoE [50]. A more practical and general HTPD workflow was developed by Welsh et al. involving a multistep approach of HT chromatography techniques as a guidance for defining the operating space [106]. No detailed modeling tools were implemented as accurate performance predictions were not the aim, only isotherm models to regress the partitioning coefficient and maximum binding capacity were used. Weigel et al. applied a similar method as Welsh et al. to investigate the effectiveness of hydrophobic interaction chromatography (HIC) as a final purification step for a cell culture-derived influenza A and B virus [107]. 96-well filter plate experiments were used for screening various resins and salt concentrations, followed by conventional lab-scale columns for dynamic binding capacity characterization. However, the major reason for choosing a rational step-wise method over mathematical modeling was the lack of available virus purification data by HIC to be able to determine model parameters. As vaccine platform processes are barely available yet, Ladd Effio et al. initiated a capture step as first part of a generic purification platform process for virus-like particles (VLP) [108]. Ladd Effio et al. established a one-step removal of HCPs and DNA from a complex VLP feedstock with an anion-exchange membrane capture step by making use of HTE and mechanistic modeling for *in silico* optimization purposes [108]. Although equilibrium and binding capacities of membrane chromatography are often limited, at high flowrates

membrane chromatography outperforms conventional packed bed chromatography in terms of productivity and for short residence times also in bed utilization [109]. It is expected that in the near future membrane materials with higher binding capacities will become available and therefore could overcome the restriction on surface area per unit volume of resin. The advancement in membrane chromatography technology is definitely interesting to the biopharmaceutical industry.

Even though chromatography is one of the main purification techniques for biopharmaceuticals and vaccines, other downstream process techniques are also HT-suited. Precipitation is a well-known technique to isolate a desired component such as a protein, DNA or virus and proven to be HT-suited [110] [22]. This separation technique depends on the physical and/or chemical interaction between the precipitating agent (e.g. calcium chloride, ammonium sulfate or PEG) with one or several of the components in which solubility is the most critical thermodynamic property [111]. Aqueous two-phase systems (ATPs) could also be an alternative to chromatography as it is based on liquid-liquid extraction employing two immiscible phases to separate components from mixtures. HT techniques in combination with statistical [112, 113] or mathematical/thermodynamic [114] models are a convenient method for characterizing these systems.

Analytics to monitor the process are just as important as the purification techniques itself. Analytics, however, remain a bottleneck during HTE, and consequently slow down experimentation considerably. Konstantinidis et al. provided a strategic assay deployment that helps selecting appropriate analytical methods, while preserving data quality [115]. Nonetheless, finding innovative ways to accelerate the analytical throughput would be of great merit.

Table 2. 2. Overview of HTPD approaches which have been applied in industry and/or academia using HT and/or modeling techniques.

Purpose of research	Experimental Method			Modeling method	Application, unit operation and stage of development	Ref
	HT	DoE	Lab-scale	Empirical or Mechanistic		
Defining characterization/ operating space	Robocolumns for condition screening	Full factorial DoE, varying two process variables, pH and buffer molarity.	None	Empirical: Regression analysis by Least square fitting and optimization using contour profiler.	- GCSF - IEX - Early	[50]
	Batch adsorption experiments for isotherms and Robocolumns breakthrough experiments (DBC)	None	Lab-scale columns, validation experiments	Mechanistic: Regression of partitioning coefficient and maximum binding capacity, Langmuir isotherm model.	- mAb - IEX, MMC - Early	[106]
Resin selection and operating conditions	Batch binding and condition screening for resin-protein interactions	None	Lab-scale columns, comparison of results between lab-scale columns and HTS filter plate results	Calculation of partition coefficient and the separation factor.	- mAb IgG1 from CHO - HIC - Early	[125]
	Batch binding and conditions screening for resin-protein interactions	None	None	Empirical: Partition coefficient of the product fitted to a response surface model (ANOVA) of pH and total chloride concentration.	- mAb IgG1 from CHO - CEX, AEX - Early	[126]
Resin selection and salt concentration	Batch adsorption experiments for condition screening	None	Lab-scale column, breakthrough experiments (DBC – 10%)	None	- Virus, influenza A, B - HIC - Early	[107]

Table 2. 3. Continuation, Overview of HTPD approaches which have been applied in industry and/or academia using HT and/or modeling techniques.

Purpose of research	Experimental Method			Modeling method	Application, unit operation and stage of development	Ref
	HT	DoE	Lab-scale	Empirical or Mechanistic		
Design bind-and-elute membrane process	Batch binding and buffer screenings	Buffer screenings	Lab-scale column, membrane characterization, breakthrough experiments	Mechanistic: General rate model for radial flow chromatography Regression and chromatogram fitting for estimating isotherm and model parameters	- Virus Like Particles - AEX, membrane chromatography - Early and late	[108]
Resin selection, optimization and defining operation window	Robocolumns for resin screening and optimization screenings	Definitive screening designs for resin screening Central composite designs for optimization screenings	Lab-scale column for model verification	Empirical: Multivariate data analysis and usage of multi-criteria decision-making techniques. Process parameter optimization and Robustness analysis	- Highly aggregate antibody solution. - CEX - Early	[127]
Resin selection, multiple unit optimization	Robocolumns, bind-elute mode, resin and operating condition screening	None	None	Empirical: Multi-objective mixed integer nonlinear programming model. Adopted ε-constraint method solved by Dinkel Bach' s algorithm	- Recombinant Fc Fusion protein - CEX - MMC - Early	[86]

Table 2. 4. Continuation, Overview of HTPD approaches which have been applied in industry and/or academia using HT and/or modeling techniques.

Purpose of research	Experimental Method			Modeling method	Application, unit operation and stage of development	Ref
	HT	DoE	Lab-scale	Empirical or Mechanistic		
Flowsheet optimization, resin selection, design of process	Batch adsorption experiments (HT) for isotherm determination and resin selection.	None	Lab-scale column experiments for validation and acquisition of molecular properties	Mechanistic: Flowsheet optimization top-to-bottom approach using chromatographic mechanistic models including adsorption isotherm models	- mAb from hybridoma cell culture - CEX, AEX, HIC, SEC - 4-steps - Early and late	[87, 90, 128]
	Robocolumns to determine isotherm parameters Batch-uptake experiments for determining maximum binding capacities	None	Lab-scale column experiments for validation	Mechanistic: Flowsheet optimization, global optimization along with ANN and Local optimization along with Mechanistic models, including isotherm models	- mAb IgG1 from CHO - CEX, MMC, HIC, UF/DF - 4-steps - Early and late	[5, 36]
	None	None	Lab-scale breakthrough column experiments.	Mechanistic: Flowsheet optimization using mechanistic models, including isotherm models.	- applied to three model proteins - CEX, AEX - 2-steps - Early and late	[25]

2.3.2. Overall purification process

The studies described in the previous paragraph focused mainly on applying HTPD to one or two sequential purification steps, but thereby do not consider the overall purification workflow. Designing a downstream process by optimizing each unit operation individually could lead to a suboptimal process design, as small variations in one-unit operation may affect the performance of subsequent following purification steps. The combination of HT and model-based optimization approaches for a sequence of unit operations has seldom been studied. Nfor et al. established a systematic approach to rationally define the protein purification process utilizing a top-to-bottom approach [90]. The least promising flowsheets were eliminated at each tree-diagram level by means of flow-sheet selection with the aim of keeping a minimum number of purification units. Instead of sequential optimization, which might generate a suboptimal process [25, 26], Huuk et al. presented a simultaneous two-step ion exchange chromatography process flowsheet optimization, including salt-gradient shapes and cut-points for fraction collection [25]. Pirrung et al. even proved the feasibility of simultaneous optimization of an integrated process consisting of three chromatographic steps (e.g. cation exchange, hydrophobic interaction and mixed-mode), including buffer exchange steps in between (e.g. ultra- and diafiltration) applied to a complex biological feedstock purification [5]. First the isotherm parameters were acquired utilizing HT techniques as illustrated in more detail in their previous work [36], hence other parameters were obtained by conventional lab-scale experiments. The use of ANN for finding suitable starting conditions for the local optimization using mechanistic models enabled circumvention of speed-limitations [5, 27]. These examples to optimize an overall downstream process require a comprehensive combination of modeling and experimental methods. If more HTPD approaches are established that combine efficiently all available technologies (e.g. LHS, modeling-, analytical-, and data-processing tools), this optimization strategy could become more interesting.

2.4. Artificial Intelligence in process development

The interest in HTPD raised after the introduction of HT technologies, having the major benefit to generate more data while consuming less material. Nevertheless, these arising technologies still face a number of hurdles. Experimentally transferring every item into HT mode, including analytics, remains a burden and although more data is being produced, processing and handling these data efficiently is still challenging. Modeling is a promising tool to close this gap. Further advancements of modeling are discussed in the following paragraph.

While complex mechanistic models attempt to describe the mechanisms and thermodynamic phenomena, determining certain parameters is rather difficult. Simplifying models could avoid certain difficulties, however, oversimplifications may cause inaccurate predictions and meaningless results. The optimal model should be as simple as possible while still gaining high or sufficient understanding. Moreover, a trade-off between accuracy versus speed has to be made especially when running optimizations. This led to the question; how to reduce the computational time effort or simplify complex models while retaining a similar level of accuracy and/or detail.

Although ANNs were already used in the late 90s to predict retention times in chemical chromatography [116, 117]. Due to the generation of larger data-sets and better computer systems in recent years, the use of AI gained popularity in various technology fields, likewise within the biotechnology area. In 1992, Psichogios and Ungar presented the first hybrid neural network-first principles approach applied to model a fed-batch bioreactor [118]. This hybrid model used a neural network model to estimate unknown process parameters serving as an input to a first principle model, resulting in an improved inter- and extrapolating capability, and understanding over merely “black-box” neural networks. Von Stosch et al. extensively reviewed the hybrid semi-parametric modeling framework, as explained in 2.2.3.3., and the various applications in (bio)chemical engineering concerning process monitoring, control, optimization, model-reduction and scale-up [61]. Nagrath et al. established an optimization framework using a serial white- and black-box configuration (Figure 2.4) to find the optimal design for a chromatographic process applied to a binary and tertiary mixtures [119]. After obtaining the physical model parameters experimentally, numerous simulations were performed under various conditions using the physical model (i.e. white-box) for training the neural network. Finally, the optimal operating conditions for several purity levels were identified by using the neural network to accelerate the computation. Likewise, Pirrung et al. used an ANN to accelerate a flowsheet optimization consisting of three chromatography and UF/DF units [5, 27]. However, here the ANN was used to find adequate starting conditions during the global optimization to be used for the local optimization, which was performed

together with a mechanistic model in order to assure realistic and accurate results. A speed improvement of 70% was found, including training of the networks, compared to using solely mechanistic models for the optimization. Reducing the computational cost was the main objective for these latter two examples (Nagrath et al. and Pirrung et al.), and therefore using ANNs was advantageous. However, the data-driven model, here ANNs, depends on the accuracy of the mechanistic model and so limits the predictive power of ANNs. Recently, Nikita et al. showed a novel approach making use of reinforcement learning (RL) to increase computational efficiency during a continuous chromatography process optimization [120]. Each mechanistic model simulation is rewarded according to a RL-method and consequently the optimization criteria (design space) are adjusted to accelerate the convergence of optimization. The optimal flowrate, directly related to yield and purity demands, was found three times faster using the RL based optimization method compared to conventional trial and error methods. However, thorough understanding of the RL-principle and mechanistic modeling is required to develop this RL-method. Apart from using hybrid semi-parametric modeling for optimization intentions, other research showed the usefulness of black-box modeling to estimate certain white-box model parameters that are hard to determine. Wang et al. used neural networks to directly derive mass transfer, isotherm and characteristic charge parameters from experimental chromatograms, after which these parameters served as input for the mechanistic model [121]. In this way, time-consuming experimental methods for determining these parameters were circumvented. However, this approach requires still a considerable number of experiments. In mechanistic filtration models the flux is a key parameter, but predicting this parameter accurately might be quite complex. Therefore, Krippl et al. used an ANN to determine the flux using transmembrane pressure, cross-flow velocity and concentration as input parameters [122]. Placing the hybrid model in series enabled to perform a multistep ahead prediction of the concentration over time. In general, data-driven models combined with white-box models can be advantageous in terms of prediction accuracy, computing and model development efficiency and enhanced extrapolation properties [61].

With an eye on the future more applications of hybrid modeling approaches are expected, in both industry and academia. In order to realize this prospective, more experts in modeling are needed to develop and maintain these software applications. Moreover, the modeling techniques utilized in the process development (HTPD) can also be used for process control and optimization in later development stages and manufacturing processes. One step ahead is industry 4.0, known as the latest revolution and aiming to digitalize the whole manufacturing process. From process control to decision-making, all monitored data is efficiently collected, which in turn is also valuable for process development [123]. In order to

realize Industry 4.0, digital twins are highly essential, defined as a virtual counterpart of the physical process and their interconnection [124].

2.5. Summary and Conclusion

Vaccination protects millions of people from infectious diseases and, because a high product quality is pivotal, the downstream processing is likewise as important. Downstream process operations in manufacturing have a direct influence on time-to-market, product quality and cost of goods. Therefore, modernizing the strategies for developing processes could be of great merit. The urge to decrease the process development timeline of vaccines has raised, as well as the need for deeper process understanding as stated by the quality by design guideline.

The introduction of high throughput technology accelerated experimental data generation and allowed to investigate the influence of parameters more thoroughly and systematically. However, HT also required to enhance data-processing and modeling techniques. Mechanistic models provide insights on the inner working mechanism of unit operations and are being increasingly adopted by industry in recent years, proving they add deeper process understanding and greater application possibilities. The combination of HT and modeling techniques led to HTPD approaches, acquiring and using data in a more efficient and purposeful way, thereby also enabling standardized process development approaches. The future direction in process development is to design and optimize the overall downstream process *in silico*, for which only a limited number of model calibration and validation experiments are needed. Hybrid (semi-parametric) modeling can help to ease the model development or improve the accuracy by making optimal use of both mechanistic and data-driven models. Recent research has shown the potential of artificial neural networks in addition to mechanistic models for circumvention of computational speed limitation or estimation of parameters.

With these emerging new technologies, it will now be possible to standardize process development workflows, provided that a proficient combination of experimenting and modeling techniques is utilized. Creating a generic process development workflow will enhance process development time and shared knowledge among different departments and manufacturing sites.

Acknowledgment

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 Vaccines and Technical University of Delft for their valuable input.

2.6. References

[1] F.E. Andre, R. Booy, H.L. Bock, J. Clemens, S.K. Datta, T.J. John, B.W. Lee, S. Lolekha, H. Peltola, T.A. Ruff, M. Santosham, H.J. Schmitt, Vaccination greatly reduces disease, disability, death and inequity worldwide, Bulletin of the World Health Organization 86(2) (2008) 140-146. https://doi.org/10.2471/BLT.07.040089.

[2] C.M.C. Rodrigues, S.A. Plotkin, Impact of Vaccines; Health, Economic and Social Perspectives, Frontiers in Microbiology 11(1526) (2020). https://doi.org/10.3389/fmicb.2020.01526.

[3] J. Ehreth, The global value of vaccination, Vaccine 21(7) (2003) 596-600. https://doi.org/https://doi.org/10.1016/S0264-410X(02)00623-0.

[4] N. Arora, Y. Al Mazrou, A. Cravioto, e. al., 2014 Assessment report of the global vacinne action plan, in: WHO (Ed.) 2014.

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

[6] E.P. Wen, R. Ellis, N.S. Pujar, Vaccine Development and Manufacturing, Wiley2014.

[7] M. Zhao, M. Vandersluis, J. Stout, U. Haupts, M. Sanders, R. Jacquemart, Affinity chromatography for vaccines manufacturing: Finally ready for prime time?, Vaccine 37(36) (2019) 5491-5503. https://doi.org/https://doi.org/10.1016/j.vaccine.2018.02.090.

[8] F. Krammer, SARS-CoV-2 vaccines in development, Nature 586(7830) (2020) 516-527. https://doi.org/10.1038/s41586-020-2798-3.

[9] P. Ball, The lightning-fast quest for COVID vaccines - and what it means for other diseases, Nature 589 (2021) 16- 18. https://doi.org/https://doi-org.tudelft.idm.oclc.org/10.1038/d41586-020-03626-1.

[10] ICH, ICH Harmonised Tripartite Guideline: Pharmaceutical Development Q8 (R2), ICH, 2009.

[11] FDA, PAT Guidance for Industry - A Framework for innovative Pharmaceutical Development, Manufacturing and Quality Assurance, 2004. www.fda.gov/regulatory-

information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance.

[12] L.X. Yu, Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control, Pharmaceutical Research 25(4) (2008) 781-791. https://doi.org/10.1007/s11095-007-9511-1.

[13] A.S. Rathore, Roadmap for implementation of quality by design (QbD) for biotechnology products, Trends Biotechnol 27(9) (2009) 546-553. https://doi.org/10.1016/j.tibtech.2009.06.006.

[14] A.S. Rathore, Quality by Design (QbD)-Based Process Development for Purification of a Biotherapeutic, Trends Biotechnol 34(5) (2016) 358-370. https://doi.org/10.1016/j.tibtech.2016.01.003.

[15] K.M. Lacki, High throughput process development in biomanufacturing, Curr Opin Chem Eng 6 (2014) 25-32. https://doi.org/10.1016/j.coche.2014.08.004.

[16] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: knowledge-based chromatographic process development, Trends Biotechnol 32(4) (2014) 210-220. https://doi.org/10.1016/j.tibtech.2014.02.001.

[17] M.N. São Pedro, T.C. Silva, R. Patil, M. Ottens, White paper on high-throughput process development for integrated continuous biomanufacturing, Biotechnology and Bioengineering n/a(n/a) (2021). https://doi.org/https://doi.org/10.1002/bit.27757.

[18] S.B. Carvalho, C. Peixoto, M.J.T. Carrondo, R.J.S. Silva, Downstream processing for influenza vaccines and candidates: An update, Biotechnology and Bioengineering n/a(n/a) (2021). https://doi.org/https://doi.org/10.1002/bit.27803.

[19] M. Jones, N. Palackal, F. Wang, G. Gaza-Bulseco, K. Hurkmans, Y. Zhao, C. Chitikila, S. Clavier, S. Liu, E. Menesale, N.S. Schonenbach, S. Sharma, P. Valax, T. Waerner, L. Zhang, T. Connolly, “HIGH-RISK” HOST CELL PROTEINS (HCPs): A MULTI-COMPANY COLLABORATIVE VIEW, Biotechnology and Bioengineering n/a(n/a) (2021). https://doi.org/https://doi.org/10.1002/bit.27808.

[20] J.O. Josefsberg, B. Buckland, Vaccine process technology, Biotechnology and Bioengineering 109(6) (2012) 1443-1460. https://doi.org/10.1002/bit.24493.

[21] A. Abdulrahman, A. Ghanem, Recent advances in chromatographic purification of plasmid DNA for gene therapy and DNA vaccines: A review, Anal Chim Acta 1025 (2018) 41-57. https://doi.org/10.1016/j.aca.2018.04.001.

[22] B. Kalbfuss-Zimmermann, U. Reichl, Viral Vaccines Purification, Vaccine Development and Manufacturing (2014) 97-180. https://doi.org/https://doi.org/10.1002/9781118870914.ch5.

[23] Y.-p. Yang, T. D'Amore, Protein Subunit Vaccine Purification, Vaccine Development and Manufacturing2014, pp. 181-216. https://doi.org/https://doi.org/10.1002/9781118870914.ch6.

[24] V. Kumar, A. Bhalla, A.S. Rathore, Design of experiments applications in bioprocessing: Concepts and approach, Biotechnology Progress 30(1) (2014) 86-99. https://doi.org/10.1002/btpr.1821.

[25] T.C. Huuk, T. Hahn, A. Osberghaus, J. Hubbuch, Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography, Sep Purif Technol 136 (2014) 207-222. https://doi.org/10.1016/j.seppur.2014.09.012.

[26] B. Otero, M. Degerman, T.B. Hansen, E.B. Hansen, B. Nilsson, Model-based design and integration of a two-step biopharmaceutical production process, Bioproc Biosyst Eng 37(10) (2014) 1989-1996. https://doi.org/10.1007/s00449-014-1174-9.

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

[28] A.A. Shukla, J. Thommes, Recent advances in large-scale production of monoclonal antibodies and related proteins, Trends Biotechnol 28(5) (2010) 253-261. https://doi.org/10.1016/j.tibtech.2010.02.001.

[29] P. Baumann, J. Hubbuch, Downstream process development strategies for effective bioprocesses: Trends, progress, and combinatorial approaches, Eng Life Sci 17(11) (2017) 1142-1158. https://doi.org/10.1002/elsc.201600033.

[30] V. Czitrom, One-Factor-at-a-Time versus Designed Experiments, The American Statistician 53(2) (1999) 126-131. https://doi.org/10.1080/00031305.1999.10474445.

[31] G.E.P. Box, D.W. Behnken, Simplex-Sum Designs: A Class of Second Order Rotatable Designs Derivable From Those of First Order, Ann. Math. Statist. 31(4) (1960) 838-864. https://doi.org/10.1214/aoms/1177705661.

[32] D.B. Hibbert, Experimental design in chromatography: A tutorial review, Journal of Chromatography B 910 (2012) 2-13. https://doi.org/https://doi.org/10.1016/j.jchromb.2012.01.020.

[33] S.L.C. Ferreira, R.E. Bruns, E.G.P. da Silva, W.N.L. dos Santos, C.M. Quintella, J.M. David, J.B. de Andrade, M.C. Breitkreitz, I.C.S.F. Jardim, B.B. Neto, Statistical designs and response surface techniques for the optimization of chromatographic systems, Journal of Chromatography A 1158(1) (2007) 2-14. https://doi.org/https://doi.org/10.1016/j.chroma.2007.03.051.

[34] J.J. Siirola, Industrial Applications of Chemical Process Synthesis, in: J.L. Anderson (Ed.), Advances in Chemical Engineering, Academic Press1996, pp. 1-62. https://doi.org/https://doi.org/10.1016/S0065-2377(08)60201-X.

[35] A.T. Hanke, E. Tsintavi, M.D.R. Vazquez, L.A.M. van der Wielen, P.D.E.M. Verhaert, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 3D-liquid chromatography as a complex mixture characterization tool for knowledge-based downstream process development, Biotechnology Progress 32(5) (2016) 1283-1291. https://doi.org/10.1002/btpr.2320.

[36] S.M. Pirrung, D.P. da Cruz, A.T. Hanke, C. Berends, R.F.W.C. van Beckhoven, M.H.M. Eppink, M. Ottens, Chromatographic Parameter Determination for Complex Biological Feedstocks, Biotechnology Progress 34(4) (2018) 1006-1018. https://doi.org/10.1002/btpr.2642.

[37] L.J. Benedini, D. Figueiredo, J. Cabrera-Crespo, V.M. Gonçalves, G.G. Silva, G. Campani, T.C. Zangirolami, F.F. Furlan, Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures, Journal of Chromatography A 1613 (2020) 460685. https://doi.org/https://doi.org/10.1016/j.chroma.2019.460685.

[38] B.K. Nfor, P.D.E.M. Verhaert, L.A.M. van der Wielen, J. Hubbuch, M. Ottens, Rational and systematic protein purification process development: the next generation, Trends Biotechnol 27(12) (2009) 673-679. https://doi.org/10.1016/j.tibtech.2009.09.002.

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

[40] J. Chen, S.M. Cramer, Protein adsorption isotherm behavior in hydrophobic interaction chromatography, Journal of Chromatography A 1165(1) (2007) 67-77. https://doi.org/https://doi.org/10.1016/j.chroma.2007.07.038.

[41] G. Carta, A. Jungbauer, Protein Chromatography: Process Development and Scale-up, 2010.

[42] M. Moreno-González, V. Girish, D. Keulen, H. Wijngaard, X. Lauteslager, G. Ferreira, M. Ottens, Recovery of sinapic acid from canola/rapeseed meal extracts by adsorption, Food Bioprod Process 120 (2020) 69-79. https://doi.org/https://doi.org/10.1016/j.fbp.2019.12.002.

[43] C.A. Brooks, S.M. Cramer, Steric mass-action ion exchange: Displacement profiles and induced salt gradients, AIChE Journal 38(12) (1992) 1969-1978. https://doi.org/10.1002/aic.690381212.

[44] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Determination of parameters for the steric mass action model—A comparison between two

approaches, Journal of Chromatography A 1233 (2012) 54-65. https://doi.org/https://doi.org/10.1016/j.chroma.2012.02.004.

[45] D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch, Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications, Biotechnology Progress n/a(n/a) (2020) e2984. https://doi.org/10.1002/btpr.2984.

[46] M. Wiendahl, P. Schulze Wierling, J. Nielsen, D. Fomsgaard Christensen, J. Krarup, A. Staby, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes – Miniaturization, Automation and Parallelization of Breakthrough and Elution Studies, Chem Eng Technol 31(6) (2008) 893-903. https://doi.org/10.1002/ceat.200800167.

[47] R. Bhambure, K. Kumar, A.S. Rathore, High-throughput process development for biopharmaceutical drug substances, Trends Biotechnol 29(3) (2011) 127-135. https://doi.org/10.1016/j.tibtech.2010.12.001.

[48] N. Singh, S. Herzer, Downstream Processing Technologies/Capturing and Final Purification, in: B. Kiss, U. Gottschalk, M. Pohlscheidt (Eds.), New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins, Springer International Publishing, Cham, 2018, pp. 115-178. https://doi.org/10.1007/10_2017_12.

[49] A.S. Rathore, D. Kumar, N. Kateja, Recent developments in chromatographic purification of biopharmaceuticals, Biotechnology Letters 40(6) (2018) 895-905. https://doi.org/10.1007/s10529-018-2552-1.

[50] R. Bhambure, A.S. Rathore, 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) (2013) 403-414. https://doi.org/https://doi.org/10.1002/btpr.1705.

[51] M. Bensch, P. Schulze Wierling, E. von Lieres, J. Hubbuch, High Throughput Screening of Chromatographic Phases for Rapid Process Development, Chem Eng Technol 28(11) (2005) 1274-1284. https://doi.org/https://doi.org/10.1002/ceat.200500153.

[52] K.M. Lacki, High-throughput process development of chromatography steps: advantages and limitations of different formats used, Biotechnol J 7(10) (2012) 1192-202. https://doi.org/10.1002/biot.201100475.

[53] T. Bergander, K.M. Lacki, High-throughput process development: Chromatography media volume definition, Eng Life Sci 16(2) (2016) 185-189. https://doi.org/https://doi.org/10.1002/elsc.201400240.

[54] M.E. Lienqueo, J.A. Asenjo, Use of expert systems for the synthesis of downstream protein processes, Comput Chem Eng 24(9) (2000) 2339-2350. https://doi.org/https://doi.org/10.1016/S0098-1354(00)00590-1.

[55] B.K. Nfor, T. Ahamed, G.W.K. van Dedem, L.A.M. van der Wielen, E.J.A.X. van de Sandt, M.H.M. Eppink, M. Ottens, Design strategies for integrated protein purification processes: challenges, progress and outlook, J Chem Technol Biot 83(2) (2008) 124-132. https://doi.org/10.1002/jctb.1815.

[56] E.W. Leser, J.A. Asenjo, Rational design of purification processes for recombinant proteins, Journal of Chromatography B: Biomedical Sciences and Applications 584(1) (1992) 43-57. https://doi.org/https://doi.org/10.1016/0378-4347(92)80008-E.

[57] L. Hagel, G. Jagschies, G. Sofer, Handbook of Process Chromatography, Development, Manufacturing, Validation and Economics, 2008.

[58] G.H.L. Sciences, Recombinant Protein Purification Handbook, Principles and Methods, 2012.

[59] A. Hagen, J. Aunins, P. DePhillips, C.B. Oswald, J.P. Hennessey Jr, J. Lewis, M. Armstrong, C. Oliver, C. Orella, B. Buckland, R. Sitrin, Development, preparation, and testing of VAQTA®, a highly purified hepatitis A vaccine, Bioprocess Engineering 23(5) (2000) 439-449. https://doi.org/10.1007/s004499900157.

[60] A.A. Shukla, B. Hubbard, T. Tressel, S. Guhan, D. Low, Downstream processing of monoclonal antibodies—Application of platform approaches, Journal of Chromatography B 848(1) (2007) 28-39. https://doi.org/https://doi.org/10.1016/j.jchromb.2006.09.026.

[61] M. von Stosch, R. Oliveira, J. Peres, S.F. de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput Chem Eng 60 (2014) 86-101. https://doi.org/10.1016/j.compchemeng.2013.08.008.

[62] P.-A. Muller, F. Fondement, B. Baudry, B. Combemale, Modeling modeling modeling, Software & Systems Modeling 11(3) (2012) 347-359. https://doi.org/10.1007/s10270-010-0172-x.

[63] J. Bezivin, O. Gerbe, Towards a precise definition of the OMG/MDA framework, Proceedings 16th Annual International Conference on Automated Software Engineering (ASE 2001), 2001, pp. 273-280.

[64] D. Bonvin, C. Georgakis, C.C. Pantelides, M. Barolo, M.A. Grover, D. Rodrigues, R. Schneider, D. Dochain, Linking Models and Experiments, Ind Eng Chem Res 55(25) (2016) 6891-6903. https://doi.org/10.1021/acs.iecr.5b04801.

[65] B. Selic, The pragmatics of model-driven development, IEEE Software 20(5) (2003) 19-25. https://doi.org/10.1109/MS.2003.1231146.

[66] L. Ljung, T. Glad, Modeling of dynamic systems, Englewood Cliffs (N.J.) : Prentice-Hall1994.

[67] A.S. Rathore, S. Mittal, M. Pathak, A. Arora, Guidance for performing multivariate data analysis of bioprocessing data: Pitfalls and recommendations, Biotechnology Progress 30(4) (2014) 967-973. https://doi.org/https://doi.org/10.1002/btpr.1922.

[68] J.P.C. Kleijnen, Response Surface Methodology, Handbook of Simulation Optimization, New York, NY : Springer New York : Springer2015, pp. 81-104. https://doi.org/10.1007/978-1-4939-1384-8_4.

[69] D. Baş, İ.H. Boyacı, Modeling and optimization I: Usability of response surface methodology, Journal of Food Engineering 78(3) (2007) 836-845. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2005.11.024.

[70] C. Anirban Roy, B. Paramita, S.P. Gandham, Development of suitable solvent system for downstream processing of biopolymer pullulan using response surface methodology, Plos One, 2013.

[71] A. Eon-Duval, K. Gumbs, C. Ellett, Precipitation of RNA impurities with high salt in a plasmid DNA purification process: Use of experimental design to determine reaction conditions, Biotechnology and Bioengineering 83(5) (2003) 544-553. https://doi.org/https://doi.org/10.1002/bit.10704.

[72] M. Toueille, A. Uzel, J.-F. Depoisier, R. Gantier, Designing new monoclonal antibody purification processes using mixed-mode chromatography sorbents, Journal of Chromatography B 879(13) (2011) 836-843. https://doi.org/https://doi.org/10.1016/j.jchromb.2011.02.047.

[73] M.-J. Chiang, M. Pagkaliwangan, S. Lute, G. Bolton, K. Brorson, M. Schofield, Validation and optimization of viral clearance in a downstream continuous chromatography setting, Biotechnology and Bioengineering 116(9) (2019) 2292-2302. https://doi.org/https://doi.org/10.1002/bit.27023.

[74] D. Solle, B. Hitzmann, C. Herwig, M. Pereira Remelhe, S. Ulonska, L. Wuerth, A. Prata, T. Steckenreiter, Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation, Chemie Ingenieur Technik 89(5) (2017) 542-561. https://doi.org/https://doi.org/10.1002/cite.201600175.

[75] S.M. Pirrung, M. Ottens, High Throughput Process Development, in: A. Staby, A.S. Rathore, S. Ahuja (Eds.), Preparative Chromatography for Separation of Proteins, John Wiley & Sons, Inc.2017, pp. 269 - 292.

[76] D.M. Ruthven, Principles of adsorption and adsorption processes, Wiley, New York, 1984.

[77] A. Felinger, G. Guiochon, Comparison of the Kinetic Models of Linear Chromatography, Chromatographia 60(1) (2004) S175-S180. https://doi.org/10.1365/s10337-004-0288-7.

[78] I. Langmuir, THE CONSTITUTION AND FUNDAMENTAL PROPERTIES OF SOLIDS AND LIQUIDS. PART I. SOLIDS, Journal of the American Chemical Society 38(11) (1916) 2221-2295. https://doi.org/10.1021/ja02268a002.

[79] B.K. Nfor, J. Ripic, A. van der Padt, M. Jacobs, M. Ottens, Model-based high-throughput process development for chromatographic whey proteins separation, Biotechnology Journal 7(10) (2012) 1221-1232. https://doi.org/10.1002/biot.201200191.

[80] L.K. Shekhawat, M. Chandak, A.S. Rathore, Mechanistic modeling of hydrophobic interaction chromatography for monoclonal antibody purification: process optimization in the quality by design paradigm, J Chem Technol Biot 92(10) (2017) 2527-2537. https://doi.org/10.1002/jctb.5324.

[81] M. Moreno-González, D. Keulen, J. Gomis-Fons, G.L. Gomez, B. Nilsson, M. Ottens, Continuous adsorption in food industry: The recovery of sinapic acid from rapeseed meal extract, Sep Purif Technol 254 (2021) 117403. https://doi.org/https://doi.org/10.1016/j.seppur.2020.117403.

[82] J. Gomis-Fons, H. Schwarz, L. Zhang, N. Andersson, B. Nilsson, A. Castan, A. Solbrand, J. Stevenson, V. Chotteau, Model-based design and control of a small-scale integrated continuous end-to-end mAb platform, Biotechnology Progress 36(4) (2020) e2995. https://doi.org/https://doi.org/10.1002/btpr.2995.

[83] K. Westerberg, N. Borg, N. Andersson, B. Nilsson, Supporting Design and Control of a Reversed-Phase Chromatography Step by Mechanistic Modeling, Chem Eng Technol 35(1) (2012) 169-175. https://doi.org/https://doi.org/10.1002/ceat.201000505.

[84] N. Andersson, A. Löfgren, M. Olofsson, A. Sellberg, B. Nilsson, P. Tiainen, Design and control of integrated chromatography column sequences, Biotechnology Progress 33(4) (2017) 923-930. https://doi.org/https://doi.org/10.1002/btpr.2434.

[85] M.M. Papathanasiou, F. Steinebach, M. Morbidelli, A. Mantalaris, E.N. Pistikopoulos, Intelligent, model-based control towards the intensification of downstream processes, Comput Chem Eng 105 (2017) 173-184. https://doi.org/https://doi.org/10.1016/j.compchemeng.2017.01.005.

[86] S. Liu, S. Gerontas, D. Gruber, R. Turner, N.J. Titchener-Hooker, L.G. Papageorgiou, Optimization-based framework for resin selection strategies in biopharmaceutical purification process development, Biotechnology progress 33(4) (2017) 1116-1126. https://doi.org/10.1002/btpr.2479.

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

[88] E.J. Close, J.R. Salm, D.G. Bracewell, E. Sorensen, A model based approach for identifying robust operating conditions for industrial chromatography with process variability, Chem Eng Sci 116 (2014) 284-295. https://doi.org/https://doi.org/10.1016/j.ces.2014.03.010.

[89] S. Vogg, T. Müller-Späth, M. Morbidelli, Design space and robustness analysis of batch and counter-current frontal chromatography processes for the removal of antibody aggregates, Journal of Chromatography A 1619 (2020) 460943. https://doi.org/https://doi.org/10.1016/j.chroma.2020.460943.

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

[91] J. Schmölder, M. Kaspereit, A Modular Framework for the Modelling and Optimization of Advanced Chromatographic Processes, Processes 8(1) (2020) 65.

[92] A. Hamidi, H. Kreeftenberg, L. van der Pol, S. Ghimire, L.A.M. van der Wielen, M. Ottens, Process Development of a New Haemophilus influenzae Type b Conjugate Vaccine and the Use of Mathematical Modeling to Identify Process Optimization Possibilities, Biotechnology Progress 32(3) (2016) 568-580. https://doi.org/10.1002/btpr.2235.

[93] A. Löfgren, M. Yamanee-Nolin, S. Tallvod, J.G. Fons, N. Andersson, B. Nilsson, Optimization of integrated chromatography sequences for purification of biopharmaceuticals, Biotechnology Progress 35(6) (2019) e2871. https://doi.org/https://doi.org/10.1002/btpr.2871.

[94] T. Hahn, T. Huuk, V. Heuveline, J. Hubbuch, Simulating and Optimizing Preparative Protein Chromatography with ChromX, Journal of Chemical Education 92(9) (2015) 1497-1502. https://doi.org/10.1021/ed500854a.

[95] S. Leweke, E. von Lieres, Chromatography Analysis and Design Toolkit (CADET), Comput Chem Eng 113 (2018) 274-294. https://doi.org/https://doi.org/10.1016/j.compchemeng.2018.02.025.

[96] K. Meyer, S. Leweke, E. von Lieres, J.K. Huusom, J. Abildskov, ChromaTech: A discontinuous Galerkin spectral element simulator for preparative liquid chromatography, Comput Chem Eng 141 (2020) 107012. https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.107012.

[97] F. Dismer, J. Hubbuch, 3D structure-based protein retention prediction for ion-exchange chromatography, Journal of Chromatography A 1217(8) (2010) 1343-1353. https://doi.org/https://doi.org/10.1016/j.chroma.2009.12.061.

[98] S. Parimal, S. Garde, S.M. Cramer, Interactions of Multimodal Ligands with Proteins: Insights into Selectivity Using Molecular Dynamics Simulations, Langmuir 31(27) (2015) 7512-7523. https://doi.org/10.1021/acs.langmuir.5b00236.

[99] S. Banerjee, S. Parimal, S.M. Cramer, A molecular modeling based method to predict elution behavior and binding patches of proteins in multimodal chromatography, Journal of Chromatography A 1511 (2017) 45-58. https://doi.org/https://doi.org/10.1016/j.chroma.2017.06.059.

[100] J. Kittelmann, K.M.H. Lang, M. Ottens, J. Hubbuch, An orientation sensitive approach in biomolecule interaction quantitative structure–activity relationship modeling and its application in ion-exchange chromatography, Journal of Chromatography A 1482 (2017) 48-56. https://doi.org/https://doi.org/10.1016/j.chroma.2016.12.065.

[101] J. Woo, S. Parimal, M.R. Brown, R. Heden, S.M. Cramer, The effect of geometrical presentation of multimodal cation-exchange ligands on selective recognition of hydrophobic regions on protein surfaces, Journal of Chromatography A 1412 (2015) 33-42. https://doi.org/https://doi.org/10.1016/j.chroma.2015.07.072.

[102] J. Kittelmann, K.M.H. Lang, M. Ottens, J. Hubbuch, Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach, Journal of Chromatography A 1510 (2017) 33-39. https://doi.org/https://doi.org/10.1016/j.chroma.2017.06.047.

[103] J.F. Buyel, J.A. Woo, S.M. Cramer, R. Fischer, The use of quantitative structure–activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production, Journal of Chromatography A 1322 (2013) 18-28. https://doi.org/https://doi.org/10.1016/j.chroma.2013.10.076.

[104] K. Rege, M. Pepsin, B. Falcon, L. Steele, M. Heng, High-throughput process development for recombinant protein purification, Biotechnology and Bioengineering 93(4) (2006) 618-630. https://doi.org/https://doi.org/10.1002/bit.20702.

[105] A. Susanto, E. Knieps-Grunhagen, E. von Lieres, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes: Assessment of Model Parameter Determination from High Throughput Compatible Data, Chem Eng Technol 31(12) (2008) 1846-1855. https://doi.org/10.1002/ceat.200800457.

[106] J.P. Welsh, M.G. Petroff, P. Rowicki, H. Bao, T. Linden, D.J. Roush, J.M. Pollard, A practical strategy for using miniature chromatography columns in a standardized high-throughput workflow for purification development of monoclonal antibodies, Biotechnology Progress 30(3) (2014) 626-635. https://doi.org/https://doi.org/10.1002/btpr.1905.

[107] T. Weigel, R. Soliman, M.W. Wolff, U. Reichl, Hydrophobic-interaction chromatography for purification of influenza A and B virus, Journal of Chromatography B 1117 (2019) 103-117. https://doi.org/https://doi.org/10.1016/j.jchromb.2019.03.037.

[108] C. Ladd Effio, T. Hahn, J. Seiler, S.A. Oelmeier, I. Asen, C. Silberer, L. Villain, J. Hubbuch, Modeling and simulation of anion-exchange membrane chromatography for purification of

Sf9 insect cell-derived virus-like particles, Journal of Chromatography A 1429 (2016) 142-154. https://doi.org/https://doi.org/10.1016/j.chroma.2015.12.006.

[109] C. Boi, A. Malavasi, R.G. Carbonell, G. Gilleskie, A direct comparison between membrane adsorber and packed column chromatography performance, Journal of Chromatography A 1612 (2020) 460629. https://doi.org/https://doi.org/10.1016/j.chroma.2019.460629.

[110] B.K. Nfor, N.N. Hylkema, K.R. Wiedhaup, P.D.E.M. Verhaert, L.A.M. van der Wielen, M. Ottens, High-throughput protein precipitation and hydrophobic interaction chromatography: Salt effects and thermodynamic interrelation, Journal of Chromatography A 1218(49) (2011) 8958-8973. https://doi.org/10.1016/j.chroma.2011.08.016.

[111] R.E. Lovrien, D. Matulis, Selective Precipitation of Proteins, Current Protocols in Protein Science 7(1) (1997) 4.5.1-4.5.36. https://doi.org/https://doi.org/10.1002/0471140864.ps0405s07.

[112] S. Zimmermann, S. Gretzinger, M.-L. Schwab, C. Scheeder, P.K. Zimmermann, S.A. Oelmeier, E. Gottwald, A. Bogsnes, M. Hansson, A. Staby, J. Hubbuch, High-throughput downstream process development for cell-based products using aqueous two-phase systems, Journal of Chromatography A 1464 (2016) 1-11. https://doi.org/https://doi.org/10.1016/j.chroma.2016.08.025.

[113] S.A. Oelmeier, F. Dismer, J. Hubbuch, Application of an aqueous two-phase systems high-throughput screening method to evaluate mAb HCP separation, Biotechnology and Bioengineering 108(1) (2011) 69-81. https://doi.org/https://doi.org/10.1002/bit.22900.

[114] B.C. Bussamra, D. Sietaram, P. Verheijen, S.I. Mussatto, A.C. da Costa, L. van der Wielen, M. Ottens, A critical assessment of the Flory-Huggins (FH) theory to predict aqueous two-phase behaviour, Sep Purif Technol 255 (2021) 117636. https://doi.org/https://doi.org/10.1016/j.seppur.2020.117636.

[115] S. Konstantinidis, E. Heldin, S. Chhatre, A. Velayudhan, N. Titchener-Hooker, Strategic assay deployment as a method for countering analytical bottlenecks in high throughput process development: Case studies in ion exchange chromatography, Biotechnology Progress 28(5) (2012) 1292-1302. https://doi.org/https://doi.org/10.1002/btpr.1591.

[116] J. Havel, J.E. Madden, P.R. Haddad, Prediction of retention times for anions in ion chromatography using Artificial Neural Networks, Chromatographia 49(9) (1999) 481. https://doi.org/10.1007/BF02467746.

[117] E. Marengo, M.C. Gennaro, S. Angelino, Neural network and experimental design to investigate the effect of five factors in ion-interaction high-performance liquid chromatography, Journal of Chromatography A 799(1) (1998) 47-55. https://doi.org/https://doi.org/10.1016/S0021-9673(97)01027-3.

[118] D.C. Psichogios, L.H. Ungar, A hybrid neural network-first principles approach to process modeling, AIChE Journal 38(10) (1992) 1499-1511. https://doi.org/10.1002/aic.690381003.

[119] D. Nagrath, A. Messac, W.B. B, M.C. S, A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes, Biotechnology Progress 20(1) (2004) 162-178. https://doi.org/10.1021/bp034026g.

[120] S. Nikita, A. Tiwari, D. Sonawat, H. Kodamana, A.S. Rathore, Reinforcement Learning based Optimization of Process Chromatography for Continuous Processing of Biopharmaceuticals, Chem Eng Sci (2020) 116171. https://doi.org/https://doi.org/10.1016/j.ces.2020.116171.

[121] G. Wang, T. Briskot, T. Hahn, P. Baumann, J. Hubbuch, Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks, Journal of Chromatography A 1487 (2017) 211-217. https://doi.org/10.1016/j.chroma.2017.01.068.

[122] M. Krippl, A. Dürauer, M. Duerkop, Hybrid modeling of cross-flow filtration: Predicting the flux evolution and duration of ultrafiltration processes, Sep Purif Technol 248 (2020) 117064. https://doi.org/https://doi.org/10.1016/j.seppur.2020.117064.

[123] J. Markarian, Industry 4.0 in Biopharmaceutical Manufacturing, Biopharm Int 31 (2018).

[124] R.M.C. Portela, C. Varsakelis, A. Richelle, N. Giannelos, J. Pence, S. Dessoy, M. von Stosch, When Is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to the Digital Twin Concept, Digital Twins, Springer Berlin Heidelberg, Berlin, Heidelberg, 2020, pp. 35-55. https://doi.org/10.1007/10_2020_138.

[125] J.F. Kramarczyk, B.D. Kelley, J.L. Coffman, High-throughput screening of chromatographic separations: II. Hydrophobic interaction, Biotechnology and Bioengineering 100(4) (2008) 707-720. https://doi.org/https://doi.org/10.1002/bit.21907.

[126] B.D. Kelley, M. Switzer, P. Bastek, J.F. Kramarczyk, K. Molnar, T. Yu, J. Coffman, High-throughput screening of chromatographic separations: IV. Ion-exchange, Biotechnology and Bioengineering 100(5) (2008) 950-963. https://doi.org/https://doi.org/10.1002/bit.21905.

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

[128] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/10.1002/bit.24576.

[129] A.J. Pollard, E.M. Bijker, A guide to vaccinology: from basic principles to new developments, Nature Reviews Immunology 21(2) (2021) 83-100. https://doi.org/10.1038/s41577-020-00479-7.

[130] R. Rappuoli, Bridging the knowledge gaps in vaccine design, Nat Biotechnol 25(12) (2007) 1361-1366. https://doi.org/10.1038/nbt1207-1361.

[131] B. Donaldson, F. Al-Barwani, V. Young, S. Scullion, V. Ward, S. Young, Virus-Like Particles, a Versatile Subunit Vaccine Platform, Subunit Vaccine Delivery (2014) 159-180. https://doi.org/10.1007/978-1-4939-1417-3_9.

Chapter 3

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

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 fifteen to only three. 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.

Published as: D. Keulen, E. van der Hagen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Using artificial neural networks to accelerate flowsheet optimization for downstream process development, Biotechnology and Bioengineering (2023) 1-14. https://doi.org/https://doi.org/10.1002/bit.28454.

3.1. 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 [1]. 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 downstream process development strategies and recent advancements has been described elsewhere [2]. DoE is based on statistical methods and most commonly applied for process development in pharmaceutical industry [3, 4]. 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 [5-7]. 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 towards 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 [8-11]. 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 [12]. The combination of AI techniques with mechanistic modeling could eliminate shortcomings in both techniques, and so improve the applicability and usability [13-15]. The potential of applying AI driven models for process development and their practical implementations have been discussed elsewhere [16, 17].

Developing a purification process requires making decisions such as type and sequential order of purification techniques, operating conditions, and costs [18, 19]. 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 [20].

Kawajiri described different optimization strategies for chromatographic modeling and summarized related studies [21]. Moreover, an open-software optimization framework for modeling conventional and advanced batches and continuous chromatography processes was developed by Schmölder and Kaspereit [22]. However, both studies are applicable to batch or continuous chromatography, but not to flowsheet optimization. Nfor et al. applied a top-to-bottom optimization approach to obtain a minimum number of purification steps in the final process [19]. As sequential optimization can lead to a suboptimal process, Huuk et al. simultaneously optimized a two-step ion-exchange chromatography process [18]. A similar approach was applied by Pirrung et al. 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) [23].

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. already established a hybrid model optimization framework for preparative chromatography, using ANNs for speed improvement [24]. In the work of Pirrung et al. 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 [23, 25]. 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. In order 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-squared 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.

3.2. Materials & Methods

3.2.1. 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

$$y = [y_1, y_2, \dots, y_n] \quad \textit{Eq. 3.1}$$

$$s.t. \quad \sum y \geq 1 \quad \textit{Eq. 3.2}$$

$$y_i \neq y_j \; for \; all \; y_i > 0 \quad \textit{Eq. 3.3}$$

$$For \; i = 2, 3, \dots, n: \quad \textit{Eq. 3.4}$$
$$if \;\; y_i > 0, then \; y_{i-1} > 0$$

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 i^{th} element of vector y. The first statement, Eq. 3.1, defines the set of all possible vectors y, where each element is number between 0 to 3, which in this study represents the considered chromatography modes, none, CEX, AEX, and HIC, respectively. The second statement assures, Eq. 3.2, to have at least one unit operation present in the sequence. The third statement, Eq. 3.3, ensures that each mode can only appear once in the sequence. The conditional constraint in Eq. 3.4 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 3.2.5. Case study. 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 3.1, 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.

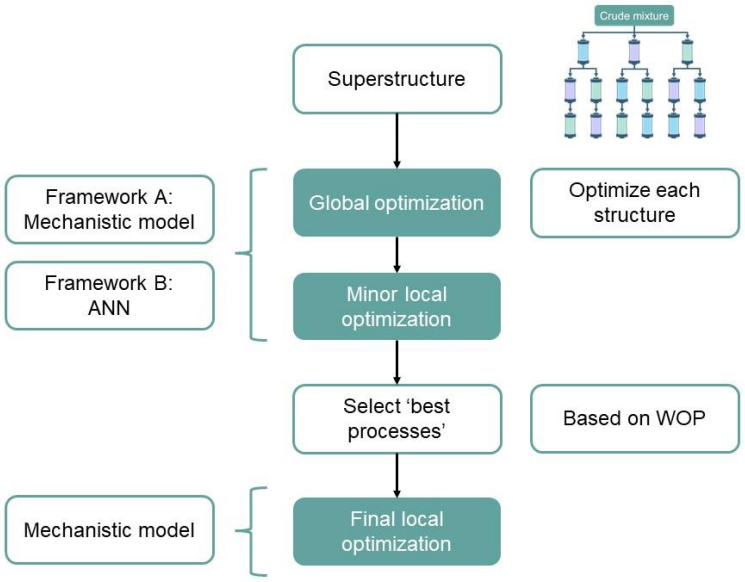


Figure 3. 1. Each flowsheet of the superstructure, upper right figure, is first optimized globally to select the 'best processes'. These are further optimized using a final local optimizer. Framework A used MMs and framework B used ANNs for global optimization.

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

$$WOP = 0.5 * purity + 0.3 * yield + 0.2 * (100 - buffer\ consumption), \quad \text{Eq. 3.5}$$

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 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.

3.2.2. 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:

$$\frac{\partial C_i}{\partial t} + F\frac{\partial q_i}{\partial t} = -u\frac{\partial C_i}{\partial x} + D_{L,i}\frac{\partial^2 C_i}{\partial x^2}, \qquad \textit{Eq. 3.6}$$

$$\frac{\partial q_i}{\partial t} = k_{ov,i}\left(C_i - C^*_{eq,i}\right), \qquad \textit{Eq. 3.7}$$

$$k_{ov,i} = \left[\frac{d_p}{6k_{f,i}} + \frac{d_p^2}{60\varepsilon_p D_{p,i}}\right]^{-1}, \qquad \textit{Eq. 3.8}$$

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 [26]. 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 [27]. Moreover, we used the multicomponent mixed-mode isotherm, as formulated by Nfor et al. [28] and described in Appendix 3.B.

3.2.3. 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 [29]. 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 [30]. 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

$$\sigma(z) = \sigma\left(\sum_{i=1}^{j} w_i \cdot X_i + b\right), \qquad \textit{Eq. 3.9}$$

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, and 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 [31].

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 3.2). In order 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.

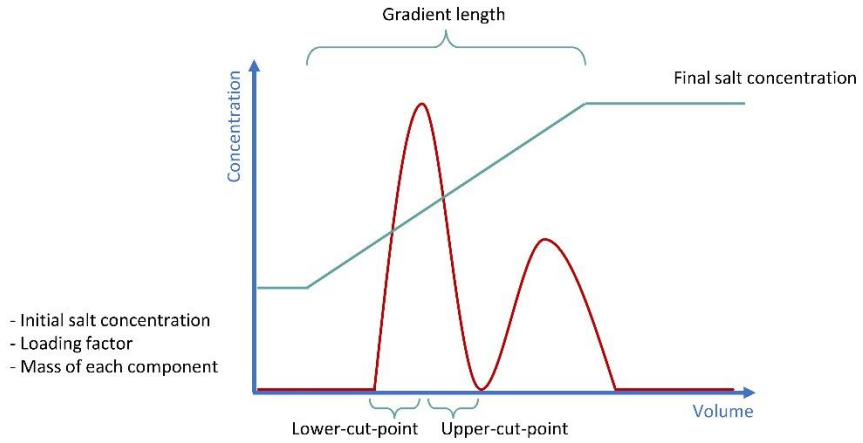


Figure 3. 2. Input parameters used for the ANN. Initial salt concentration ranging from 5−300 mM for CEX and AEX, and 100−500 mM for HIC. The final salt concentration ranging from 100−1200 mM for CEX and AEX, and 5−300 mM for HIC. The gradient length in the range of 1−10 CV, and loading factor from 0.05−5 CV. The concentration, converted to mass, ranging from 0.001−4 g/L. Both cut points in percentage of the peak maximum, lower cut point from 1−80% and upper cut points from 20−99%.

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 [32]. 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. Numerical methods. The used hyperparameters for each ANN of each chromatography mode are given in Table 3.1.

Table 3. 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

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:

$$R^2 = 1 - \frac{\sum_i^N (y_i - Y_i)^2}{\sum_i^N (Y_i - \bar{Y}_i)^2}, \qquad \textit{Eq. 3.10}$$

$$RMSE = \sqrt{\frac{\sum_i^N (y_i - Y_i)^2}{N}}, \qquad \textit{Eq. 3.11}$$

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.

3.2.4. 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 3.A.

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 Differentia Equations) algorithm from the scipy.integrate package. This method automatically switches between the nonstiff Adams method and the stiff BDF method [33].

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

1–80% of the peak maximum, and for the upper cut point between 20–99% of the peak maximum. The initial salt was between 5–300 mM, and the final salt between 100–1200 mM for CEX and AEX. For HIC the initial salt was between 100–500 mM and the final salt concentration between 5–300 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, 10C, and a 8GB Nvidia Quadro of 8GB. Multiple cores were used to execute the simulations most efficiently, however the number of cores varied depending on the simulation.

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.

3.2.5. 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 to 5) were considered, data was taken from a previous study [34]. The protein names can be found in Appendix 3.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 3.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

$$min\, f(x) = (100 - yield\,(x)) + 2 * (100 - purity\,(x)) + eluent\ consumption(x) \quad \textit{Eq. 3.12}$$

$$s.t.\quad h(x) = 0 \qquad (only\ applies\ to\ MM) \quad \textit{Eq. 3.13}$$

$$0 \leq x \leq 1, \quad \textit{Eq. 3.14}$$

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 (Eq. 3.14). Moreover, when using the MM, the equality equations $h(x)$ must be complied, which are mass balances and equilibrium relations (Eq. 3.13). 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.

3.3. Results & Discussion

3.3.1. 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 3.2. 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 ($4.81 \cdot 10^{-8}$ - 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 3.3 (A, B) for CEX). Ten ANNs were trained for each number of sample points, and the unseen test data was 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 assume 15 flowsheets have to be evaluated 1000 times, this will result in a total of about 33.000 simulations with the MM [25]. 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 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 3.3 (A, 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 3.3 (C – F) for CEX. This overall evaluation for each chromatography mode and each protein can be found in Appendix 3.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 3.1.

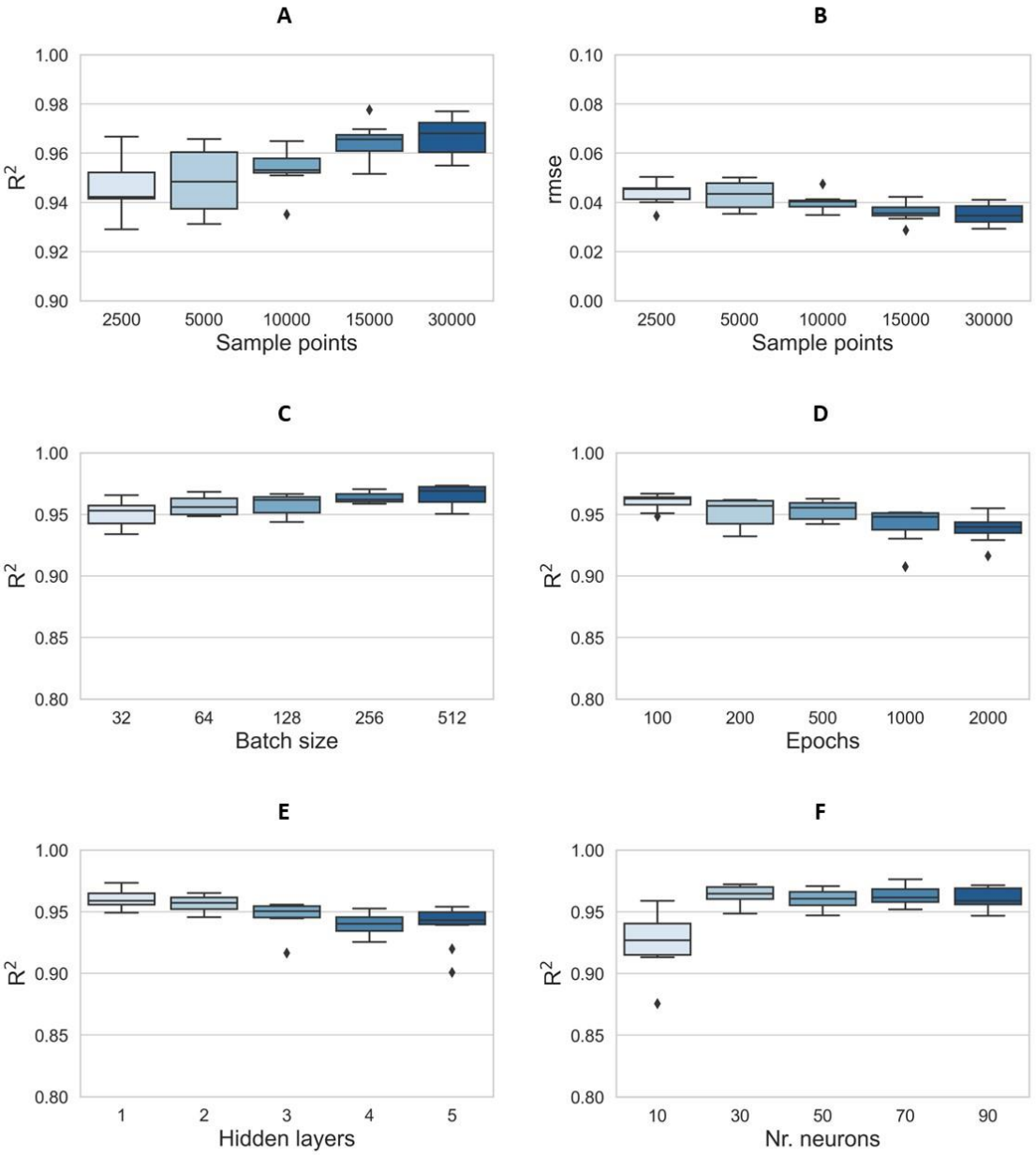


Figure 3. 3. Boxplots A & B show the accuracy (left: R^2 and right: RMSE value) for different number of sample points for the CEX chromatography mode using the data of the product. Boxplots C – F show the effect of varying certain hyperparameters on the R^2 for protein 1 in CEX chromatography. The standard hyperparameters were 3 hidden layers each having 50 neurons, a batch size of 128 and epoch of 200, the number of sample points used was 10.000.

Table 3. 2. Quantitative evaluation for each chromatography mode and all proteins. 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.

		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

The quantitative evaluation of each ANN is shown in Table 3.2, most of the ANNs reached the desired values of $R^2 > 0.90$ and RMSE <0.04. The generated data is 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 ($<1 \cdot 10^{-5}$). 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 3.4, for proteins 1, 4, and 5 during CEX for unseen test data. The data points are aligned close to the 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 3.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 3.4 and Appendix 3.D.

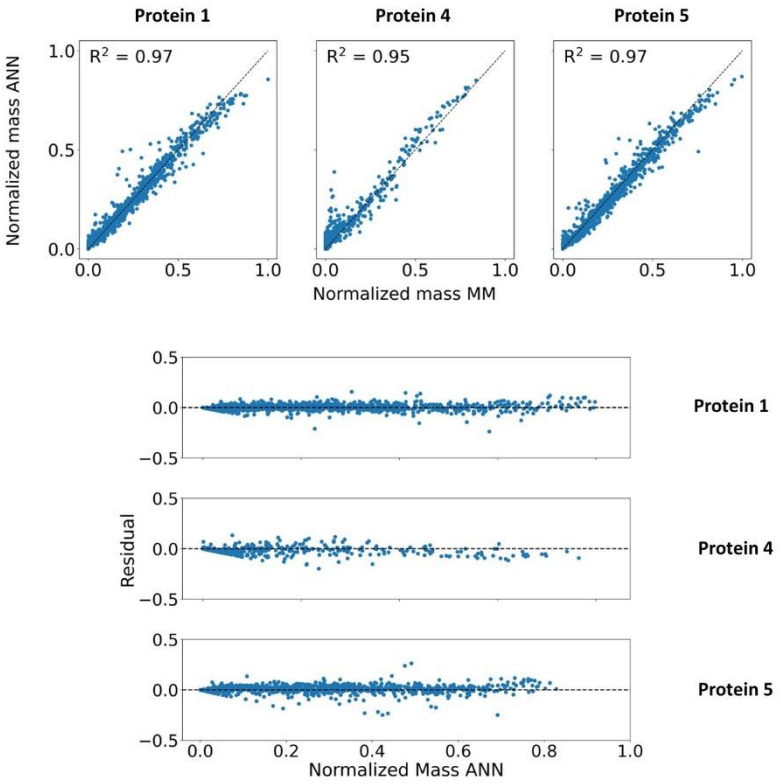


Figure 3. 4. Upper figure: Prediction capabilities for the normalized ANN outcome of mass against the outcome of MM. Lower figure: Residuals showing the difference between predicted mass values by the ANN and the MM. Both plots show unseen test-data (1493 points) for the proteins 1, 4, and 5 for the CEX mode.

Contour plots for each chromatography mode were made to qualitatively evaluate the ANNs, Figure 3.5. 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 3.5 (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 3.5 (1a)). The overprediction by ANNs is due to the standard deviation and results in an overprediction of the yield, e.g., mass output divided by the mass injected.

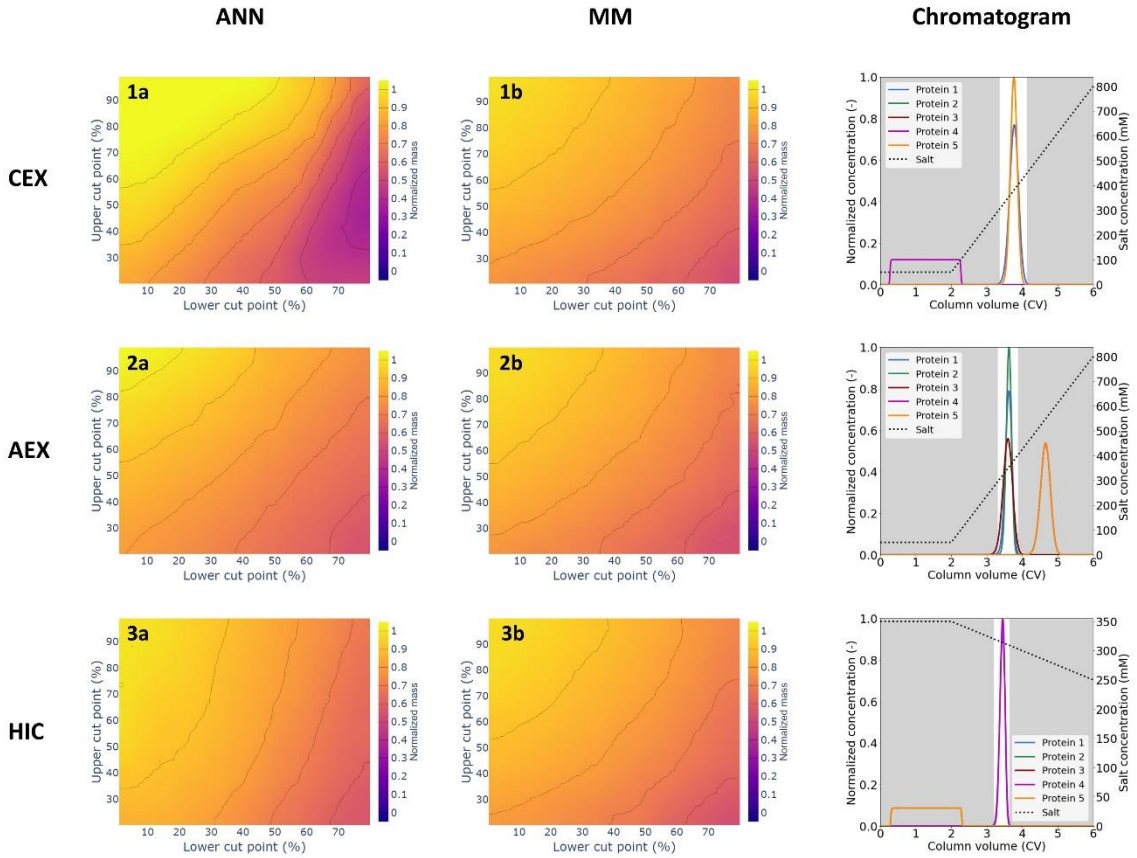


Figure 3. 5. Qualitative analysis of the ANNs compared to the true values of the MMs for each chromatography mode (CEX, AEX, and HIC). In the first column the predicted results of the ANN for a varying range of cut points is shown, in the second column the outcome of the MM is shown. In the last column, 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.

3.3.2. Flowsheet optimization

Optimizing a flowsheet is a multimodal optimization where multiple global optima could be present [35]. 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 details can be found in section 3.2.4. 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

3.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 3.2.1. Flowsheet optimization workflow). The performance results of the global optimization using either MMs or ANNs is shown in Table 3.3.

Table 3. 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	53.2	99.8	100	10.62	4.07	0.85	1.8	75	76
1	CEX - AEX	99.99	79.7	99.3	94.9	3.32	1.4	2.02	5.4	99	87
2	CEX - AEX - HIC	99.97	75	97.5	89.3	1.57	1.49	5.09	6.7	98	83
3	CEX - HIC	100	60	98.9	100	13.51	1.39	3.33	5	99	79
4	CEX - HIC - AEX	99.72	75.7	97.4	100	0.94	1.94	4.54	6.6	98	87
5	AEX	49.81	49.7	99.5	100	1.31	1.92	2.51	2.9	74	74
6	AEX - CEX	99.92	99.8	99.2	100	1.47	1.18	2.88	3.9	99	99
7	AEX - CEX - HIC	99.87	99.8	97.7	84	1.02	1.36	5.01	9.9	98	93
8	AEX - HIC	99.93	99.7	98.6	100	33.15	1.39	3.11	3.9	99	99
9	AEX - HIC - CEX	99.99	99.8	98.7	100	1.48	1.41	4.58	5	99	99
10	HIC	49.98	63.1	99.6	100	36.05	90.8	0.8	0.8	75	81
11	HIC - CEX	100	99.7	98.5	100	3.55	11.7	3.18	3.3	99	99
12	HIC - CEX - AEX	99.99	97.6	97	89.6	3.59	1	4.9	3.8	98	95
13	HIC - AEX	100	99	99.4	100	5.32	16.9	3.03	1.6	99	99
14	HIC - AEX - CEX	99.92	98.6	97.9	86.6	1.66	0.7	3.57	7.7	99	94

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 a sufficient purity. As seen in Figure 3.5, the extreme cut points were overpredicted 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 3.2). 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 3.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 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. [19]. The performance results differ because other process conditions, objective, and variables were applied. Also, Pirrung et al. performed a similar study in which the optimal found sequence was CEX – HIC, in this study equal sequence 3 [25]. 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., when using MMs during he optimization [25]. 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 3.6. The range of the initial salt concentration varied between 5–300 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 3.6 (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.

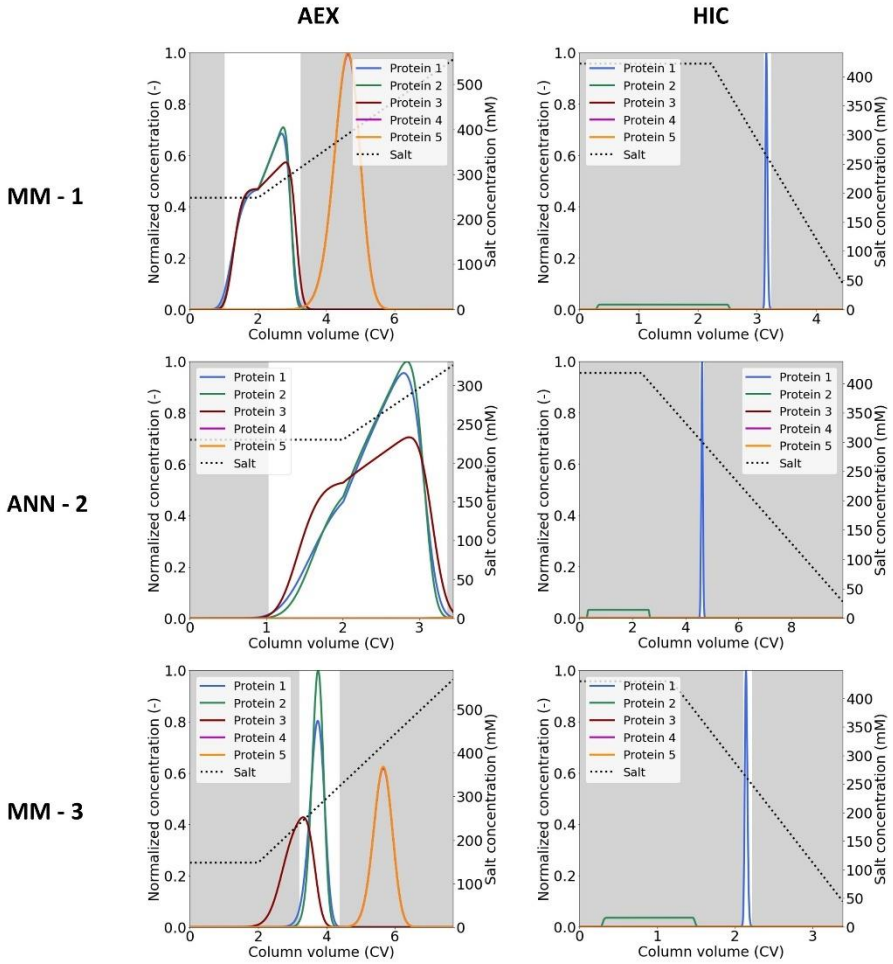


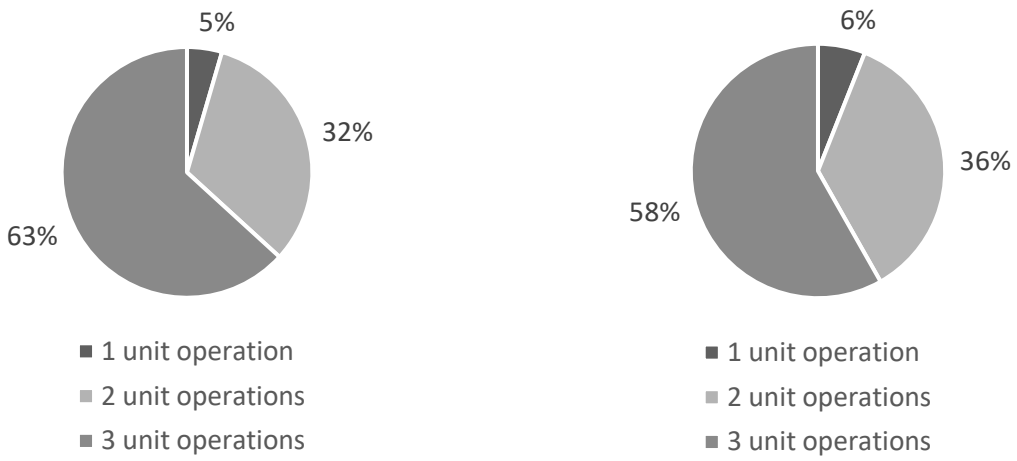
Figure 3. 6. Outcome of the final local optimization using either the global outcome of the MM (first row) or the ANN (second row) as a starting condition. The third row (MM – 3) shows the final local optimization outcome if the initial salt concentration is adjusted to a maximum of 150 mM. Protein 1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

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 3.4. As expected for ANNs, about 97% of the total simulation time is spend 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 3.4. 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 hours for the mechanistic model and 0.08 hours 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.

Table 3. 4. Comparison of time spend for the global optimization when using MMs or ANNs. The data generation for the ANNs includes simulations of 10.000 sample points for each chromatography mode and the training of the 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



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.

3.4. 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 hours, while using ANNs took about half of the time so 200 hours.

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 application 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.

Acknowledgment

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.

3.5. References

[1] K.M. Łącki, Chapter 16 - Introduction to Preparative Protein Chromatography, in: G. Jagschies, E. Lindskog, K. Łącki, P. Galliher (Eds.), Biopharmaceutical Processing, Elsevier2018, pp. 319-366. https://doi.org/https://doi.org/10.1016/B978-0-08-100623-8.00016-5.

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

[3] R. Bhambure, A.S. Rathore, 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) (2013) 403-414. https://doi.org/https://doi.org/10.1002/btpr.1705.

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

[5] ICH, ICH Harmonised Tripartite Guideline: Pharmaceutical Development Q8 (R2), ICH, 2009.

[6] FDA, PAT Guidance for Industry - A Framework for innovative Pharmaceutical Development, Manufacturing and Quality Assurance, 2004. www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance.

[7] L.X. Yu, Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control, Pharmaceutical Research 25(4) (2008) 781-791. https://doi.org/https://doi.org/10.1007/s11095-007-9511-1.

[8] F. Silva, D. Resende, M. Amorim, M. Borges, 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) (2020) 175. https://doi.org/https://doi.org/10.3390/joitmc6040175.

[9] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review, Processes 8(9) (2020) 1088. https://doi.org/https://doi.org/10.3390/pr8091088.

[10] I.C. Reinhardt, D.J.C. Oliveira, D.D.T. Ring, Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector, Journal of Industrial Information Integration 18 (2020) 100131. https://doi.org/https://doi.org/10.1016/j.jii.2020.100131.

[11] R.M.C. Portela, C. Varsakelis, A. Richelle, N. Giannelos, J. Pence, S. Dessoy, M. von Stosch, When Is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to

the Digital Twin Concept, Digital Twins, Springer Berlin Heidelberg, Berlin, Heidelberg, 2020, pp. 35-55. https://doi.org/https://doi.org/10.1007/10_2020_138.

[12] A. Felinger, G. Guiochon, Comparison of the Kinetic Models of Linear Chromatography, Chromatographia 60(1) (2004) S175-S180. https://doi.org/https://doi.org/10.1365/s10337-004-0288-7.

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

[14] M. von Stosch, R. Oliveira, J. Peres, S.F. de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput Chem Eng 60 (2014) 86-101. https://doi.org/https://doi.org/10.1016/j.compchemeng.2013.08.008.

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

[16] M. von Stosch, R.M.C. Portela, C. Varsakelis, A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice, Curr Opin Chem Eng 33 (2021) 100692. https://doi.org/https://doi.org/10.1016/j.coche.2021.100692.

[17] A.S. Rathore, S. Nikita, G. Thakur, S. Mishra, Artificial intelligence and machine learning applications in biopharmaceutical manufacturing, Trends Biotechnol (2022). https://doi.org/https://doi.org/10.1016/j.tibtech.2022.08.007.

[18] T.C. Huuk, T. Hahn, A. Osberghaus, J. Hubbuch, Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography, Sep Purif Technol 136 (2014) 207-222. https://doi.org/https://doi.org/10.1016/j.seppur.2014.09.012.

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

[20] H. Yeomans, I.E. Grossmann, A systematic modeling framework of superstructure optimization in process synthesis, Comput Chem Eng 23(6) (1999) 709-731. https://doi.org/https://doi.org/10.1016/S0098-1354(99)00003-4.

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

[22] J. Schmölder, M. Kaspereit, A Modular Framework for the Modelling and Optimization of Advanced Chromatographic Processes, Processes 8(1) (2020) 65. https://doi.org/https://doi.org/10.3390/pr8010065.

[23] S.M. Pirrung, C. Berends, A.H. Backx, R.F.W.C. van Beckhoven, M.H.M. Eppink, M. Ottens, Model-based optimization of integrated purification sequences for biopharmaceuticals,

Chemical Engineering Science: X 3 (2019) 100025. https://doi.org/https://doi.org/10.1016/j.cesx.2019.100025.

[24] D. Nagrath, A. Messac, W.B. B, M.C. S, A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes, Biotechnology Progress 20(1) (2004) 162-178. https://doi.org/https://doi.org/10.1021/bp034026g.

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

[26] D.M. Ruthven, Principles of adsorption and adsorption processes, John Wiley & Sons, New York, 1984.

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

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

[29] J.E. Madden, N. Avdalovic, P.R. Haddad, J. Havel, 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) (2001) 173-179. https://doi.org/https://doi.org/10.1016/S0021-9673(00)01185-7.

[30] A.C. Müller, S. Guido, Introduction to machine learning with python : a guide for data scientists O'Reilly Media2017.

[31] C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: Comparison of trends in practice and research for deep learning, arXiv preprint arXiv:1811.03378 (2018).

[32] M. Fellner, A. Delgado, T. Becker, Functional nodes in dynamic neural networks for bioprocess modelling, Bioproc Biosyst Eng 25(5) (2003) 263-270. https://doi.org/https://doi.org/10.1007/s00449-002-0297-6.

[33] L. Petzold, Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations, SIAM Journal on Scientific and Statistical Computing 4(1) (1983) 136-148. https://doi.org/https://doi.org/10.1137/0904010.

[34] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/https://doi.org/10.1002/bit.24576.

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

Chapter 4

Comparing *in silico* flowsheet optimization strategies in biopharmaceutical downstream processes

The challenging task of designing biopharmaceutical downstream processes is initially to select the type of unit operations, followed by optimizing their operating conditions. For complex flowsheet optimizations, the strategy becomes crucial in terms of duration and outcome. In this study, we compared three optimization strategies, namely, simultaneous, top-to-bottom, and superstructure decomposition. Moreover, all strategies were evaluated by either using chromatographic Mechanistic Models (MMs) or Artificial Neural Networks (ANNs). An overall evaluation of 39 flowsheets was performed, including a buffer-exchange step between the chromatography operations. All strategies identified orthogonal structures to be optimal, and the weighted overall performance values were generally consistent between the MMs and ANNs. In terms of time-efficiency, the decomposition method with MMs stands out when utilizing multiple cores on a multiprocessing system for simulations. This study analyses the influence of different optimization strategies on flowsheet optimization and advices on suitable strategies and modeling techniques for specific scenarios.

This chapter has been submitted for publication: Keulen, D., Apostolidi, M., Geldhof, G., Le Bussy, O., Pabst, M., and Ottens, M..

4.1. Introduction

Downstream processing is of major importance for delivering the required quality and quantity of a biopharmaceutical product, which has to meet the strict standards by regulatory authorities [1]. The downstream process is a substantial expense of the overall manufacturing costs, therefore, an efficient and cost-effective process is crucial. One of the major, most costly, and essential purification techniques is chromatography, which is capable to achieve very high product purities [2]. Eventually, the combination of purification steps will determine the overall process performance. Therefore, developing a purification process is a challenging task, involving many variables, such as type and sequential order of purification techniques, operating conditions, and costs [3, 4]. A comprehensive overview of the different strategies in downstream process development together with the latest breakthroughs was given recently by Keulen et al. [5]. Finding an optimal purification process at an early stage of the process design is desirable in terms of costs, quality, and development time. Flowsheet optimization evaluates all process possibilities in-silico, which can support the decision-making for an early process design. For many years, flowsheet optimization has been applied to design chemical processes, therefore, it is well-known in the field of process systems engineering [6, 7].

Around the 1970s, the first articles were published about process design synthesis [8, 9]. Sirrola et al. developed a general computer-aided process synthesizer that was able to select process equipment and the system configurations [8]. Umeda et al. presented an integrated optimization approach to optimize two alternative routes for a distillation system [9]. Over the past five decades, the field of superstructure-based optimizations has evolved greatly, along with the intensified computing possibilities [10]. Mencarelli et al. provides an adequate overview of superstructure-based optimization history, superstructure representation types, and modeling strategies [6]. Most superstructure-based optimizations applied in chemical engineering are related to reactor networks [11], distillation processes [12], and heat exchangers [13]. Several programs are available to perform a chemical superstructure-based optimization, for example, P-graph [14], Pyosyn [15], and Super-O [16]. As most of these chemical process simulations are based on first-principle models this can be computationally time-consuming, therefore the interest in employing surrogate models for optimization purposes increased. In 1998, Altissimi et al. already showed the value of replacing a first-principle model with a surrogate model for optimization purposes [17]. Afterwards, more research followed on using surrogate or meta-models for superstructure or complex optimization purposes [18-22].

Despite the biopharmaceutical industry only emerged about 40 years ago, this industry is advancing rapidly and shifting towards Industry 4.0 [23-25]. Industry 4.0 desires to entirely

digitalize the manufacturing process, aiming to implement and combine model-based process development techniques with efficiently stored monitored data. Hence, realizing the utilization of Digital Twins, which are digital models of the real process and enable to directly control the real process [26, 27]. In this way, more knowledge can be acquired about the processes, which is in compliance with the Quality by Design guidelines [28, 29]. A general biopharmaceutical process consists of an upstream and downstream part, in which the downstream part focuses on the purification of the biopharmaceutical. The purification steps can be subdivided into capture, intermediate, and polishing steps as shown in Figure 4.1. The main purpose of the capture and intermediate steps is to concentrate, isolate, and stabilize the product, and remove the majority of the impurities. While the subsequent polishing steps target high purity values [2].

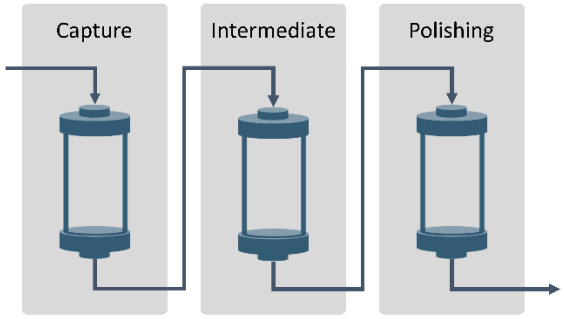


Figure 4. 1. Simplified schematic overview of the chromatography steps in a biopharmaceutical downstream process, the sequence can also have less or more chromatography operations depending on the process. The capture step aims to concentrate, isolate, and stabilize the product, together with the Intermediate step, their target is to remove the bulk impurities. The main purpose of the polishing step is to attain high product purities.

Chromatographic MMs have been around for several years, and industry is gradually adopting these methods [30, 31]. Lately, advances have been made to faster and more efficiently determine the adsorption isotherms, which are needed as input parameters for the mechanistic model [32-34]. Likewise, several research has been published to determine adsorption isotherms for complex mixtures [35-37]. And more recently, Disela et al. characterized the host cell proteome of two universal E. coli strains based on mass spectrometry data, which approach can be used for initial decision-making on process development [38]. Not only the techniques and methods to determine the adsorption isotherm are making progress, also the MMs are advancing in terms of speed and accuracy. Meyer et al. applied a computational more efficient method for the spatial discretization and obtained a speed-improvement of at least 20 times, for higher precision it even improves over 100 times compared to the open-software CADET [39, 40]. Their chromatography model was recently extended by Breuer et al., which applied a similar method to the particle mass

balance [41]. Rao et al. developed a 3D model to simulate the chromatography process with very high precision to acquire knowledge about the complex transport mechanism [42]. Moreover, hybrid modeling, using artificial intelligence (AI) in combination with mechanistic modeling, can overcome certain limitations of both modeling techniques [43, 44]. Narayanan et al. employed artificial neural networks (ANNs) for fitting the solid-phase mass balance, which reduced the model complexity, and an improved accuracy compared to the conventional mechanistic model was observed [45]. Accordingly, this progress in experimentally determining model-parameters, improving the MMs, and making use of hybrid modeling, is advantageous for digitalization of the downstream process and likewise for optimization purposes.

As described previously, flowsheet optimization enables to screen the overall design space and finding the optimal purification process at an early development stage. Process systems engineering recognized the added value of superstructure-based optimization for chemical processes. Also for biochemical processes, it is essential to optimize the integrated processing steps to discover the most optimal process globally [46]. Liu and Papageorgiou developed a data-driven optimization framework to find the best process according to economical and certain performance objectives [47]. However, the data for each optional processing steps is already provided and not generated internally. This type of optimization is known as biopharmaceutical manufacturing process optimization, usually based on mixed integer programming techniques [48-51]. Though, these optimizations do not use detailed mechanistic modeling techniques, they are either data-driven or using surrogate models to represent the unit operations. In the work of Nfor et al., a top-to-bottom optimization approach is performed that evaluates the performance of each unit operation at each level and disregards the least promising options [3]. As the influence of sequential steps is not incorporated in this approach, it might overlook the most promising sequence(s). Therefore, Huuk et al. performed an integrated two-step ion-exchange chromatography optimization [4]. Subsequently, Pirrung et al. performed a flowsheet optimization having a maximum of three chromatography steps (e.g., cation exchange, hydrophobic interaction, and mixed-mode) including a buffer exchange if needed, and simultaneously optimizing each flowsheet [52]. In their work, ANNs functioned as surrogate model for the MMs during the global optimization to find starting conditions for the local optimization, and so reducing the overall optimization time. However, the ANNs were infrequently able to find realistic results and for the subsequent local optimization the MMs were used, which was the most time-consuming part of the overall optimization [52, 53]. In our previous work, we extended this method by including the mass of each component as a variable and using more data to increase the ANN accuracy [54]. Subsequently, we compared ANNs, functioning as surrogate models, versus

MMs for flowsheet optimization to select the 'most promising sequences' during the global optimization. Only the 'most promising sequences' were further optimized through local optimization using MMs. The ANNs selected three out of four best flowsheets and reduced the overall computational time by 50%. However, for more complex flowsheet optimizations (e.g., including more unit operations and/or larger sequences) or when considering more components, not only the modelling technique (e.g., MMs or surrogate models) matters, but also the optimization strategy might play a significant role in the overall flowsheet optimization. Hence, what optimization strategy is most useful in terms of outcome, complexity, and time-efficiency?

In this chapter, we compared three different optimization strategies: simultaneous optimization, top-to-bottom approach, and superstructure decomposition, to evaluate which strategy would be most beneficial in terms of outcome, complexity, and time-efficiency when performing a complex flowsheet optimization. Simultaneous optimization involves optimizing all parameters simultaneously, top-to-bottom approach optimizes parameters sequentially from the initial to the final unit operation, and decomposition of the superstructure involves breaking down the process into smaller parts and optimizing each part separately. These strategies were chosen based on the difference in number of unit operations being optimized simultaneously and so the overall considered possibilities within the design space as indicated in Figure 4.2. For example, the top-to-bottom approach might overlook promising sequences, as it lacks a focus on optimizing the connections between chromatography steps. Additionally, for each optimization strategy the MMs and the ANNs are employed to evaluate their performance on a more complex optimization. In this complex flowsheet optimization, we included an optional buffer exchange between the chromatography steps, described by a filtration MM. This gives a total combination of 39 flowsheets to be evaluated.

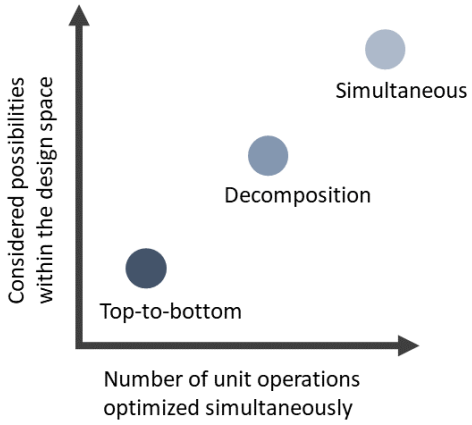


Figure 4. 2. Visualization of the difference between the chosen optimization strategies; top-to-bottom, superstructure decomposition, and the simultaneous strategy. The x-axis shows the number of unit operations being optimized simultaneously during flowsheet optimization. While, the y-axis correspondingly shows that more options in the design space are explored when the connection between chromatography steps is also considered, which is not taken into account for the top-to-bottom approach as it individually optimizes each chromatography step.

4.2. Materials & Methods

4.2.1. Flowsheet optimization workflow

First, the superstructure was generated considering a maximum of three chromatography steps and a dilution or buffer exchange by Tangential Flow Filtration (TFF) between the chromatography operations. This gives a maximum sequence of five unit operations, and at least one unit operation is needed for the purification. To generate this superstructure, confirming the defined conditions, the mathematical problem is formulated as

$$y = [y_1, y_2, \dots, y_n] \qquad \textit{Eq. 4.15}$$

$$s.t. \quad \sum y \geq 1 \qquad \textit{Eq. 4.2}$$

$$For\ i\ is\ odd: \qquad \textit{Eq. 4.3}$$
$$y_i = 1, 2, 3$$
$$y_i \neq y_{i+2}\ for\ all\ y_i > 0$$

$$For\ i\ is\ even: \qquad \textit{Eq. 4.4}$$
$$y_i = 4, 5$$

$$For\ i = 2, 3, \dots, n: \qquad \textit{Eq. 4.5}$$
$$if\ \ y_i > 0, then\ y_{i-1} > 0\ ,$$

where y is the process configuration, in which n, in this case $n = 5$, is the length of the vector. The variable $y_i \in \{0, 1, 2, 3, 4, 5\}$ represents the value of the i^{th} element of vector y. The first statement, Eq. 4.1, defines the set of all possible vectors y, where each element is an integer

number between 0 to 5, which in this study represents the considered unit operations; none, CEX, AEX, HIC, dilution, and filtration, respectively. The second statement, Eq. 4.2, guarantees that the sequence includes at least one unit operation. The third and fourth statements, Eq. 4.3 and Eq. 4.4, specify that only at odd positions in the sequence, a chromatography step is present, while for even number positions, either a dilution or filtration step is employed. Furthermore, statement three ensures that each chromatography mode appears only once in the sequence. The conditional constraint in Eq. 4.4 is applicable to all positions in the sequence, except the first position. It enforces that any occupied position in the sequence must be preceded by another occupied position. This guarantees that there are no isolated modes in the sequence and requires all modes to be connected.

The flowsheets consisting of a filtration operation to perform the buffer exchange step are modelled as a nested optimization, which means that the outer optimization involves matching the chromatography steps with their respective variables, while the inner optimization focuses on optimizing the filtration step [55]. So, for each evaluation of the outer optimization, the filtration step is always optimized internally. As the filtration model is less complex and described by ordinary differential equations (ODEs) with respect to time, it has a significantly shorter solving time compared to the chromatography model. The same flowsheet optimization workflow, as presented in our previous paper [54], was applied as shown in Figure 4.3. First, a global and minor local optimization was performed according to certain objective(s) and constraint(s), these are described in 4.2.5. Case study. For this part, either MMs or ANNs were used for the chromatography steps. After this global and minor local optimization, the most promising sequences were selected based on the weighted overall performance (WOP), which is described as follows:

$$WOP = 0.5 * purity + 0.3 * yield + 0.2 * (100 - buffer\ consumption), \quad \textit{Eq. 4.6}$$

where the calculation of purity (%) involves dividing the product amount by the total amount of proteins present in the product-pool. The yield (%) is determined by the total amount of product recovered divided by the loaded amount of product. The buffer consumption typically ranges from 1 to 50 ($L/g_{product}$). Subtracting this buffer consumption from 100 aligns it with the purity and yield ranges, and ensures that higher WOP values indicate less buffer consumption.

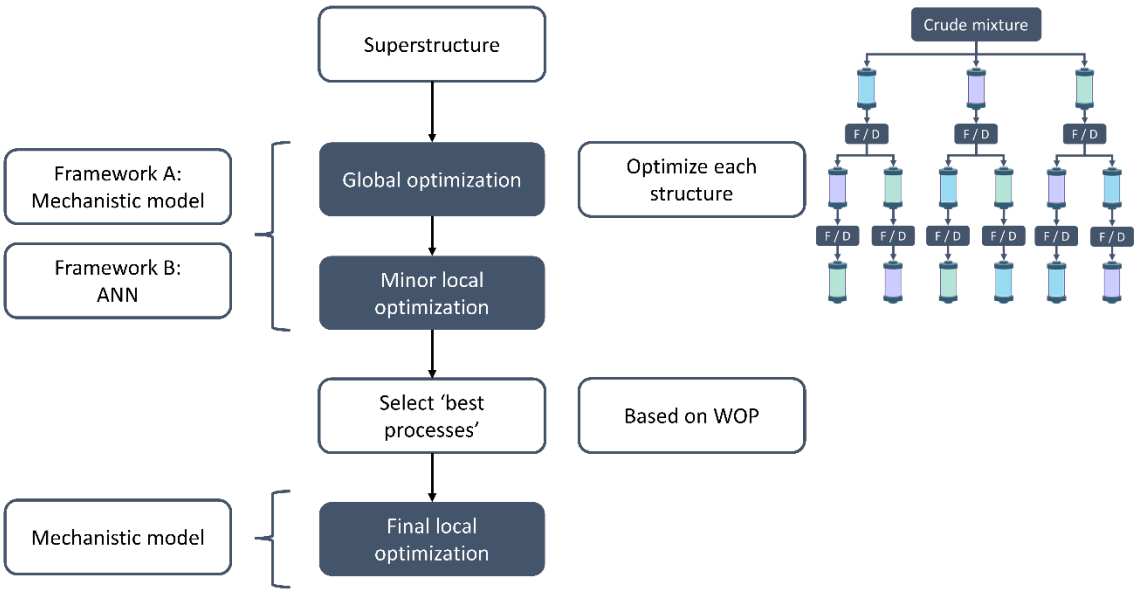


Figure 4. 3. Within the superstructure, as depicted in the upper right figure, each flowsheet is initially optimized globally to identify the most optimal processes. F / D indicates the option to have either a filtration (F) or dilution step (D). Subsequently, these selected processes are finetuned using a final local optimization step. For the global optimization, Framework A uses MMs and Framework B uses ANNs.

The selected processes were further locally optimized using the simultaneous strategy with MMs, the outcome of preceding minor local optimization was used as initial guess for the final local optimization. This flowsheet optimization workflow was applied to all three optimization strategies, the difference is the manner of solving the superstructure. Each strategy was evaluated for using either the MMs or ANNs for the global and minor local optimization. The strategies (e.g., simultaneous optimization, top-to-bottom approach, and decomposition of the superstructure) are separately described in the following sections.

4.2.1.1. Strategy I: Simultaneous flowsheet optimization

The simultaneous flowsheet optimization is the same as applied in Keulen et al. [54]. In this strategy, all parameters are optimized simultaneously, which means that the total number of variables linearly increases with the number of chromatography steps present in the sequence, as shown in Figure 4.4. For example, if five optimization variables are considered and the sequence consists of two chromatography steps, ten variables have to be optimized in total. For three unit operations, this will lead to 15 variables to be optimized.

Figure 4. 4. Schematic representation of each optimization strategy indicating with grey planes which unit operations are optimized simultaneously during the optimization. F / D indicates the option to have either a filtration (F) or dilution step (D).

4.2.1.2. Strategy II: Top-to-bottom approach

The top-to-bottom approach, based on the work of Nfor et al. [3], evaluates the superstructure by each level, see Figure 4.4. The first level optimizes the first unit operation individually. After optimizing the first level, the initial constraint assesses whether the optimal process has been achieved (e.g., purity > 99% and yield > 95%). The second constraint assesses if the optimized unit operation satisfies the minimal requirements (e.g., purity > 20% and yield > 40%) to continue to the next level, otherwise the flowsheets, starting with this type of unit operation, will be disregarded. All options present in the second level are also optimized individually. Subsequently, the overall sequence of two chromatography steps, including the dilution or the filtration operation, are simulated. The outcome of these flowsheets is evaluated by the previously described constraints, however, the second constraint is only satisfied if the purity and yield are higher than 40%. If the optimal process has not been identified yet, the optimization will continue to the third level, which operates in the same manner as the second level. If the optimal performance is not achieved after three levels, the best out of all these evaluated flowsheets can still be chosen, as all outcomes are stored. The constraints between the levels can be easily adapted to a different number and/or different performance measurements to be assessed.

4.2.1.3. Strategy III: Superstructure decomposition

In the previous study, we observed that approximately 60% of the total optimization time, whether employing MMs or ANNs, was dedicated to optimizing sequences of three unit operations [54]. This aligns with the fact that the maximum number of function evaluations increases with the number of variables to be optimized [56]. Accordingly, the question raised; does the third unit operation have a significant impact on the previous unit operations? Followed by, is it really necessary to optimize the whole process simultaneously or can we decompose the superstructure when optimizing larger sequences? In chemical engineering, different formats of decomposing the superstructure have been applied [6, 57-59]. In this study, this strategy is a combination of the simultaneous and top-to-bottom approach as shown in Figure 4.4. The superstructure is ordered in such a way that the sequences consisting of the same first three unit operations are sequential in order of length. The sequence consisting of three unit operations is optimized first, subsequently, the outcome of the third unit operation (e.g., second chromatography step) is used as input for the last chromatography step, which is optimized individually. After individually optimizing the third chromatography step, the overall sequence of five unit operations is simulated, similar to the workflow of top-to-bottom approach. In this way, only a maximum of two chromatography steps is optimized simultaneously, making the overall optimization more time-efficient compared to simultaneous optimization.

4.2.2. Chromatography

4.2.2.1. Mechanistic model

The same chromatographic MM from previous work was used [54]. The equilibrium transport dispersive model in combination with the linear driving force described the dynamic adsorption behavior during the chromatographic separation process as

$$\frac{\partial C_i}{\partial t} + F\frac{\partial q_i}{\partial t} = -u\frac{\partial C_i}{\partial x} + D_{L,i}\frac{\partial^2 C_i}{\partial x^2}, \qquad \textit{Eq. 4.7}$$

$$\frac{\partial q_i}{\partial t} = k_{ov,i}\left(C_i - C^*_{eq,i}\right), \qquad \textit{Eq. 4.8}$$

$$k_{ov,i} = \left[\frac{d_p}{6k_{f,i}} + \frac{d_p^2}{60\varepsilon_p D_{p,i}}\right]^{-1}, \qquad \textit{Eq. 4.9}$$

where C_i is the concentration in the liquid phase, q_i the concentration in the solid phase, and $C^*_{eq,i}$ is the liquid phase concentration in equilibrium with the solid phase. The phase ratio, F, is defined as $F = (1-\varepsilon_b)/\varepsilon_b$, where ε_b is the bed porosity. u represents the interstitial velocity of the mobile phase and D_L is the axial dispersion coefficient. Time and space are indicated by t and x 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 [60]. 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 [61]. Moreover, we used the linear multicomponent mixed-mode isotherm, as formulated by Nfor et al. [62] and described in Appendix 3.B. The input parameters used in this chapter are given in Appendix 4.A and 4.2.5. Case study.

4.2.2.2. Artificial Neural Networks

The ANNs were created as described previously [54]. In this work, we applied the same input variables (e.g., mass of each component, amount of loading in column volume (CV), gradient length, initial and final salt concentrations, and the lower and upper cut points in percentage of the peak maximum) and output variables (e.g., mass of each component, volume, salt concentration and each cut point in CV, salt concentration). The parameter space was based on prior-knowledge of biopharmaceutical downstream processes [63]. The data consisted of 10.000 sample points divided into 70% for training, 15% for validation, and 15% for testing.

Based on previous work, the same hyperparameters were used as starting point for developing the ANNs. Out of ten trained, validated, and tested ANNs, the best one was chosen based on R2 and root mean squared error (RMSE) values. An overview of the final used hyperparameters and applied parameter space is given in Table 4.1.

Table 4. 1. Overview of hyperparameters for each chromatography mode and the applied parameter space.

	CEX	AEX	HIC
Hyperparameters			
Batch size	512	512	512
Epochs	200	500	500
Number of hidden layers	2	2	2
Number of neurons	50	50	50
Learning rate	0.01	0.01	0.01
Parameter space			
Gradient length (CV)	1 – 10	1 – 10	1 – 10
Loading factor (CV)	0.05 – 5	0.05 – 5	0.05 – 5
Mass (g)	2e-5 – 0.39	2e-5 – 0.39	2e-5 – 0.39
Initial salt concentration (mM)	1 – 200	1 – 200	350 – 500
Final salt concentration (mM)	100 – 1200	100 – 1200	5 – 200
Lower cut point (%)	1 – 80	1 – 80	1 – 80
Upper cut point (%)	20 – 99	20 – 99	20 – 99

4.2.3. Filtration mathematical model

An ultrafiltration / diafiltration (UF/DF) mathematical model was developed to describe the buffer exchange if a filtration step was used between the chromatography steps. This model consists of first-order differential equations involving the feed solution volume (V) and added diluent volume (V_w) over time, and the solute concentrations (C_i) and the salt concentration (C_s) over time [64]. The system of mass balances for which the proteins are completely retained by the membrane is written as follows:

$$\frac{dV}{dt} = (\alpha - 1)JA, \quad \text{Eq. 4.10}$$

$$\frac{dV_w}{dt} = \alpha JA, \quad \text{Eq. 4.11}$$

$$\frac{dC_i}{dt} = \frac{C_i}{V}(\sigma_i - \alpha)JA, \quad \text{Eq. 4.12}$$

$$\frac{dC_s}{dt} = \frac{C_s}{V}(\sigma_s - \alpha)JA, \quad \text{Eq. 4.13}$$

where J is the permeate flux and A is the membrane area. σ_i and σ_s are the rejection coefficients, in this work all proteins were significantly larger than the membrane pores, hence σ_i was equal to one. While the salts could flow through and therefore σ_s was equal to zero. α is the ratio between the diluent flowrate (u) and the permeate flowrate and given as

$$\alpha = \frac{u}{JA}. \qquad \textit{Eq. 4.14}$$

The operation was performed in an ultrafiltration with variable volume diafiltration (UFVVD), therefore α can range between 0 and 1. A value close to zero indicates to operate in an UF mode, while close to one DF occurs. The flux was defined by the osmotic pressure model as

$$J = \frac{\Delta P_{TM} - \Delta\pi}{\mu * R_m}, \qquad \textit{Eq. 4.15}$$

where Δπ denotes the osmotic pressure difference and μ is the solution viscosity. ΔP_{TM} is the transmembrane pressure, which denotes the pressure difference between both sides of the membranes and acts as the driving force for the flux through the membrane. In the osmotic pressure model, the solute wall concentration is considered as a variable and increases usually over time, therefore the osmotic pressure changes, which directly impacts the flux negatively over time. The initial solute wall concentration, $C_{i,w,0}$, is predicted by solving the following equation:

$$k_0 \, ln\frac{C_{i,w,0}}{C_{i,0}} = \frac{\Delta P_{TM} - \Delta\pi}{\mu \, R_m}, \qquad \textit{Eq. 4.16}$$

where $C_{i,0}$ represent the initial concentrations in solution and k_0 is the initial mass transfer coefficient. The change of the wall concentration over time was included in the mass balance systems as

$$\frac{dC_{i,w}}{dt} = \frac{\frac{k_0}{C_i} - \ln\frac{C_{i,w}}{C_i}\frac{dk}{dC_i}}{\frac{k_0}{C_{i,w}} + \frac{1}{\mu \, R_m}\frac{\Delta\pi}{dC_{i,w}}} \frac{dC_i}{dt}, \qquad \textit{Eq. 4.17}$$

where the change of osmotic pressure is found by differentiating Eq. 4.17 with respect to $C_{i,w}$ [64]. Similarly, differentiating the mass transfer to C_i gives dk/dc_i. The mass transfer coefficient is viscosity dependent and given as follows [64]:

$$k = k_0 \left(\frac{\mu}{\mu_0}\right)^{-\frac{1}{6}}, \qquad \textit{Eq. 4.18}$$

where μ is the solution viscosity and μ_0 is the viscosity of the pure solvent. In Appendix 4.B, additional information is provided on the transmembrane pressure, osmotic pressure, second virial coefficient (B_{22}), the mass transfer correlations, and determination of the initial membrane resistance through a water flux wet experiment. Moreover, the filtration model was validated for an UF/DF wet experiment using a Bovine Serum Albumin (BSA) solution, more information can also be found in Appendix 4.B.

4.2.4. Numerical methods

The same numerical methods as applied in previous work were used, only minor adjustments were made [54]. All codes are written in Python (version 3.8). An overview of the Python libraries used is provided in Appendix 3.A. The computations were performed on a Dell Precision 5820 Tower XCTO having a 3.7G Intel Xeon processor of 3.7 GHz, 10C, and a 8GB Nvidia Quadro. Multiple cores were used to execute the simulations most efficiently; however, the number of cores varied depending on the simulation.

Dynamic chromatography column model

The Method of Lines is applied for the spatial discretization, using a fourth-order central difference scheme for both first and second-order derivatives with respect to space, to transfer partial differential equations into ODEs with respect to time. The LSODA (Livermore Solver for Ordinary Differentia Equations) algorithm from the *scipy.integrate* package is used to solve the ODEs, this method automatically switches between the nonstiff Adams method and the stiff BDF method [65].

Optimization

The *scipy.optimize* package was employed for the optimization, whereas the *differential_evolution* algorithm was used for the global optimization and Nelder-Mead algorithm for the local optimization. For global optimization, the maximum number of iterations was 6 and a population size of 5 for MMs, while for ANNs, the maximum number of iterations was 8 with a population size of 8. Latin hypercube sampling was used to initialize the population. The initial local optimization had a maximum of 5 iterations. The relative and function tolerances for both global and local optimizations were set to 1e-2. The final local optimization allowed a maximum of 50 iterations. Limited ANN accuracy can lead to varied mass predictions and affect the performance measurements. Overpredicted masses were set to the injected mass. The lower cut point ranged from 1–80% of the peak maximum, while the upper cut point ranged from 20–99% of the peak maximum. Initial salt concentrations were between 1–150 mM for CEX and AEX, 100–500 mM for HIC using MM, and 350–500 mM for HIC using ANN. Final salt concentrations were between 160–1200 mM for CEX and AEX, 5–300 mM for HIC using MM, and 5–200 mM mM for ANN. The gradient length varied from 1 to 10 CV. For optimizing the filtration operation, the Nelder-Mead algorithm with standard settings was employed.

Artificial Neural Networks

The Keras Module (version 2.10.0) of TensorFlow (version 2.10.1) were used to create the ANNs, these are open-source libraries compatible with the Python programming language.

The ANN structure, optimized with a learning rate of 0.01 using *keras.optimizers.Adam* and defined using *keras.models.Model*, employed data scaling via the *sklearn.preprocessing.MinMaxScaler* module. The optimizer loss function used the 'mean_squared_error' metric. Randomized data was generated by applying the Latin hypercube sampling method from the *pyDOE* package.

4.2.5. Case study

The case study focused on a monoclonal antibody product of interest and referred to as protein 1, and eight impurities (referred to as proteins 2 to 9), using data from a prior study [35] and additional artificial data as shown in Table 4.2. No data was available for BSA on HIC-resin, based on performed column gradient experiments, we estimated the isotherm parameters to be equal to protein 3 (Chitotriosidase), as both proteins elute at the end of the gradient. More details can be found in Appendix 4.A, as well as details about the resin parameters. The artificial data ensured at least three chromatography modes were required to purify the product of interest. Accordingly, a comprehensive comparison between the different optimization strategies could be accomplished. The chromatography column size (20.1 mL) was set in compliance to the size of the filtration unit operation. The linear flowrate of the chromatography process was 150 cm/h and the loading factor was 2.0 CV.

Table 4. 2. Input parameters of each protein used for the flowsheet optimization, in which K_{eq} is the equilibrium constant, v is the stoichiometric coefficient of salt counter ions (characteristic charge), and n is the hydrophobic interaction stoichiometric coefficient. Protein 1 = monoclonal antibody, protein 2 = Moesin, protein 3 = Chitotriosidase, protein 4 = Legumain, protein 5 = Thioredoxin reductase, protein 6 = Bovine Serum Albumin, protein 7 – 9 = artificial proteins.

Protein			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Initial concentration		g/L	1.50	0.90	0.80	1.20	1.20	0.80	0.30	0.90	1.40
Molecular weight		kDa	145.60	68.00	51.50	56.20	54.50	56.20	70.00	60.00	90.00
CEX	$\boldsymbol{K_{eq}}$	$(-)$	8.50	500.80	604.20	0.00	8.50	8.50	8.50	15.00	8.50
	$\boldsymbol{v}$	$(-)$	2.60	2.50	2.60	0.00	2.60	2.60	3.00	2.50	3.00
AEX	$\boldsymbol{K_{eq}}$	$(-)$	0.50	0.50	0.90	3.90	3.90	3.90	0.50	0.50	2.50
	$\boldsymbol{v}$	$(-)$	4.00	4.00	1.70	2.90	2.90	2.90	4.00	4.00	3.00
HIC	$\boldsymbol{K_{eq}}$	$(-)$	9.30	1.60	10.40	9.30	1.60	10.40	3.00	9.30	9.30
	$\boldsymbol{v}$	$(-)$	9.30	1.60	10.40	9.30	1.60	10.40	3.00	9.30	9.30

The validated filtration model for BSA was used to make valid assumptions for the simulation of other proteins during the overall flowsheet optimization. All proteins had similar or higher molecular weights compared to BSA, therefore full retention by the membrane was assumed for all proteins. The yield of the filtration operation was set to 95% for compensation of the lost material by adding an additional unit operation. The same constants for determining the B_{22} value, as given in Appendix 4.B, were assumed for the other proteins. However, due to the low protein concentrations evaluated in this case study, the B22 has no significant influence on the DF operation. Here, a DF mode (α=0.99) was employed to exchange buffers, e.g., adapt salt conditions, between the chromatography steps. Therefore, only the time is a variable and the optimization problem was formulated as

$$\min f(t) = \left|C_{s,model}(t) - C_{s,desired}\right| \quad \textit{Eq. 4.19}$$

$$s.t. \quad V(t_0) = V_0\ ;\ C_i(t_0) = C_{i,0}\ ;\ C_s(t_0) = C_{s,0}, \quad \textit{Eq. 4.20}$$

where t is the time variable to be optimized. $C_{s,model}$ is the model-predicted final salt concentration to be equalized to the desired final salt concentration, $C_{s,desired}$. The desired final salt concentration is in this case the initial salt concentration of the next chromatography operation.

For the flowsheet optimization, the global and local objective were formulated as

$$min\, f(x) = (\,100 - yield\,(x)\,) + 2 * (\,100 - purity\,(x)) + eluent\ consumption(x) \quad \textit{Eq. 4.21}$$

$$s.t. \quad h(x) = 0 \qquad (only\ applies\ to\ MM) \quad \textit{Eq. 4.22}$$

$$0 \le x \le 1, \quad \textit{Eq. 4.23}$$

where $f(x)$ is the objective function to be minimized, all variables (x) were normalized between 0 and 1 for enhanced optimization purposes (Eq. 4.23). Additionally applicable when using MMs is to satisfy the equality equations $h(x)$, such as the mass balances and equilibrium relations (Eq. 4.22). The optimizing variables (x) for the chromatography steps were: the gradient elution length, initial and final salt concentrations, and the lower and upper cut points. The performance measurements (e.g., yield, purity, buffer consumption) were evaluated across the entire purification process, with purity being assigned twice the weight due to its critical importance in biopharmaceutical purifications. Minimizing buffer consumption indirectly addresses the costs, batch throughput, and productivity concerns. The cost of lost feed is related to yield. Finally, the selected optimal flowsheets and their conditions from the global and minor local optimization were used as input for the final local optimization.

For both the global and local optimizers the following requirements were applied:

- Evaluation of the subsequent unit operation is only performed if the prior unit operation exceeds a yield of 5%, preventing solver failure due to excessively low concentration values.
- If the product pool's salt concentration is larger than the initial salt concentration of the next unit operation, either a dilution or filtration step is performed, depending on the flowsheet being evaluated.
- If the product pool's salt concentration is smaller than the initial salt concentration of the next unit operation, a spiking dilution step using a salt stock concentration of 5 M is performed.
- When using ANNs, the loading factor should be within the range of 0.05 and 5 CV to ensure compatibility with the data range for which the ANNs were developed. Otherwise, this option is indicated with not-a-number (Nan).

4.3. Results & Discussion

4.3.1. Filtration model validation

The filtration model was validated for the UF/DF experiment of BSA as shown in Figure 4.5. A good agreement between the experimental protein concentration and the model was found, R^2 = 0.99 and a low standard deviation of 0.03. Also the salt reduction over time is accurately predicted, R^2 = 0.97 and a standard deviation of 6.25. The alpha parameter was fitted to be 0.405, instead of the initial determined 0.7, as the permeate flowrate appeared to be not entirely constant throughout the process.

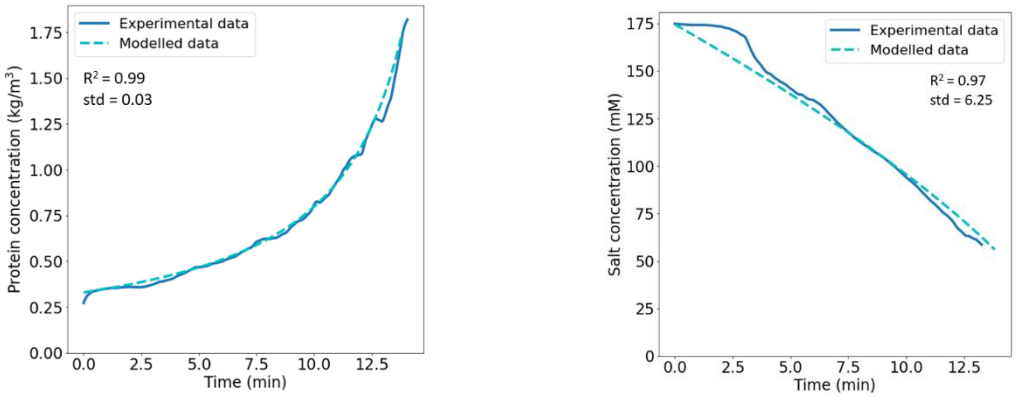


Figure 4. 5. Left: model prediction of the protein concentration, containing BSA, over time compared to the experimental values. Right: model prediction of the salt concentration over time compared to the experimental values. The initial protein concentration was 0.3 kg/m³, the initial salt concentration contained 175 mM NaCl. The initial volume was 100 mL, the flowrate was 20 mL/min. The transmembrane pressure was 0.142 MPa.

4.3.2. Artificial Neural Networks

The quantitative evaluations showed that the desired values of $R^2 > 0.90$ and RMSE < 0.04, based on previous research [54], were reached for almost all ANNs (Table 4.3). Converting the normalized RMSE values into absolute RMSE values gives an error value between 9.3–14.1% for protein 1, and for the volume between 3.6–11%. As justified in previous research, we considered an error rate of 15% to be acceptable, and to confidently identify the most optimal flowsheets while disregarding the less promising ones during flowsheet optimization. The generated data is focused around the product peak, resulting in some proteins that never elute or appear in the product pool. Therefore, training accurate ANNs is challenging due to their consistently low output values, inducing low R^2 values. Nevertheless, the absolute RMSE values also remain low ($<8 \cdot 10^{-5}$). Given our understanding that these proteins will never be present in the product pool, we can assume they would always be removed. The most challenging proteins to remove are the ones eluting around the product peak, and therefore these are considered as the critical proteins for that chromatography mode. For AEX these are the proteins: 2, 3, 7, and 8, while for CEX these are the proteins: 5, 6, 7, 8, and 9, and for HIC the proteins: 4, 8, and 9.

Table 4. 3. Quantitative evaluation for all proteins and volume on each chromatography mode. The RMSE is given as a normalized number. The product pool volume and salt concentration are included as these are needed for connecting the unit operations and calculating certain performance measurements.

	AEX		**CEX**		**HIC**	
	R^2	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Protein 1	0.99	0.016	0.99	0.020	0.98	0.022
Protein 2	0.99	0.020	-0.10	0.028	0.00	0.005
Protein 3	0.94	0.028	0.00	0.052	-0.14	0.328
Protein 4	-0.41	0.018	0.61	0.021	0.98	0.024
Protein 5	-1.08	0.014	0.99	0.021	0.00	0.010
Protein 6	-1.11	0.006	0.99	0.023	0.03	0.327
Protein 7	0.99	0.020	0.98	0.026	0.03	0.019
Protein 8	0.99	0.017	0.93	0.021	0.98	0.025
Protein 9	0.55	0.006	0.98	0.024	0.97	0.029
Volume	0.93	0.052	0.94	0.042	0.89	0.035
Salt	0.98	0.018	0.98	0.02	0.97	0.022

4.3.3. Flowsheet optimization

The flowsheet optimization workflow is designed to initially identify the global optima for each flowsheet. Subsequently, the most promising candidates can be further optimized locally, while the less promising ones may be disregarded. In this way, the number of flowsheets to be evaluated locally can be drastically reduced and correspondingly decreasing the overall

optimization time. Optimizing a complex flowsheet involves finding global optima, therefore, a stochastic and heuristic algorithm was employed to increase the chance of finding most of the global optima [66].

We compared three optimization strategies, namely, simultaneous, top-to-bottom, and decomposition, in terms of time-efficiency, complexity, and final results. Each optimization strategy was executed following the optimization workflow, as described in 4.2.1. Flowsheet optimization workflow, by either using MMs or ANNs. The flowsheet optimization was performed for a superstructure of three chromatography modes with a dilution or a filtration operation between the chromatography steps to function as a buffer exchange. In total, 39 flowsheets were evaluated. The maximum number of iterations using MMs was reduced compared to previous work to perform the flowsheet optimization within a reasonable amount of time, details can be found in 4.2.4. Numerical methods [54]. Similarly for ANNs, the number of iterations was adapted to guarantee a fair comparison between both workflows. The overall performance of each flowsheet is evaluated using the WOP value as described in 4.2.1. Flowsheet optimization workflow. In this work, the WOP is determined by the purity, yield, and buffer consumption. Based on the highest WOP value for all strategies using MMs, two best flowsheets were selected for which both MM and ANN results are shown in Table 4.4. All results of the global optimized flowsheet for all strategies, using MMs or ANNs, can be found in Appendix 4.D. Note, when the salt concentration in the pool is lower than the initial salt concentration of the subsequent chromatography step, a dilution with a stock salt solution is performed, as described in 4.2.5. Case study. This also applies to flowsheets positioned with a filtration step, and can be confirmed by evaluating the optimized variables for the salt conditions. Moreover, in the top-to-bottom strategy using ANNs, Nan occurred when the loading factor of a second or third chromatography step was out-of-range for the ANNs, as stated in the requirements in 4.2.5. Case study.

Table 4. 4. Performance measurement results of the global and minor local optimization results for the selected two best flowsheets from the MM modeling workflow, the ANN results are also provided. The selected best flowsheets for each strategy are highlighted.

Structure	Strategy	Purity (%)		Yield (%)		Buffer consumption (L/g)		WOP	
		MM	ANN	MM	ANN	MM	ANN	MM	ANN
CEX - D - HIC - D - AEX	Simultaneous	99.7	98.9	96	98.1	8.59	7.29	97*	97
	Top-to-bottom	92.7	89.6	92.3	100	7.63	4.2	93	94
	Decomposition	99.2	90.2	81.4	95.4	5.02	6.7	93*	92*
AEX - D - HIC - D - CEX	Simultaneous	99.9	99.6	89.3	100	6.48	7.23	95**	98
	Top-to-bottom	99.3	98.3	95.7	100	5.77	8.7	97	97
	Decomposition	99.1	97.3	96.2	100	6.04	10.8	97	97

*The filtration is a spiking dilution step as explained in 4.2.5. Case study for *Flowsheet [1-5-3-4-2] in Appendix table 4.D.1, 4.D.3, and 4.D.6 and for ** Flowsheet [2-5-3-4-1] in Appendix table 4.D.1.*

The strategies top-to-bottom and decomposition found the same best flowsheet (AEX – D – HIC – D – CEX), while the simultaneous strategy found a different one (CEX – D – HIC – D – AEX), as highlighted in Table 4.4. The flowsheet (AEX – D – HIC – D – CEX) was selected as an optimal candidate in all strategies when using ANNs. In overall, the ANNs found more optimal flowsheets (WOP>96) compared to MM results. This is mainly appointed to an overestimation of the yield, which depends on the ANN accuracies for each protein (Appendix 4.D). The Swarmplot, in Figure 4.6, shows the WOP values for the structures of one, two, or three chromatography steps in a sequence by either using MMs or ANNs. The different strategy outcomes are merged into the number of chromatography steps. Moreover, we clearly observe the same increasing trend when considering more chromatography steps for both ANNs and MMs. For one and two chromatography steps, the WOP value is a bit overestimated by the ANNs, mainly due to the overestimation of the yield as pointed out previously. The range for WOP values of three chromatography steps is about equal, only more flowsheets were estimated with a higher WOP value when using ANNs.

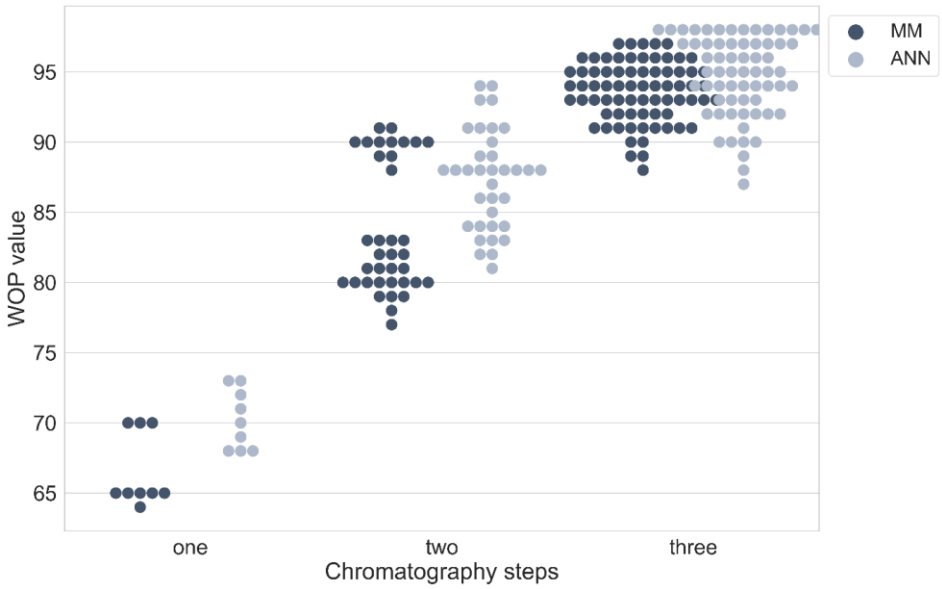


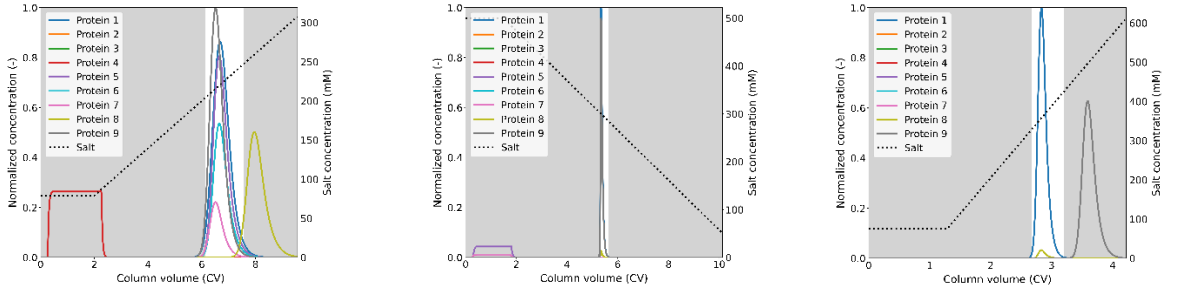
Figure 4. 6. The WOP value of each flowsheet determined by each optimization strategy is compared for one, two, and three chromatography steps, and between using either MMs or ANNs as modeling workflow.

The selected best flowsheets, for each optimization strategy with MMs, were further locally optimized using the simultaneous strategy with MMs, as shown in Figure 4.7. Noticeably, the solver objective is to discover the ideal salt conditions within sequential chromatography steps, thereby eliminating the need for filtration and so obtaining enhanced yields and reducing buffer consumptions. Often, an orthogonal structure is applied in industrial processes, meaning that ion exchange and hydrophobic interaction chromatography are

alternated [2]. Here, the two selected best flowsheets also have an orthogonal structure. However, from the global optimization results, other promising sequences, with a WOP>96, are not necessarily orthogonal. For the final local optimization, a maximum number of 50 iterations was set to minimize the computational time, which took about eight hours. From the final results in Figure 4.7, it can be observed that there is a clear trade-off between purity and yield, for example the purity result of the simultaneous strategy is reduced, while the yield increases, when comparing to the global optimized results. The buffer consumption was reduced in all strategies, but the overall WOP value was not improved for all strategies. So, to really improve the outcome, more iterations are needed. Or if a certain performance measurement, such as the purity, is a severe constraint (>99%), this can be applied to only local or both global and local optimization.

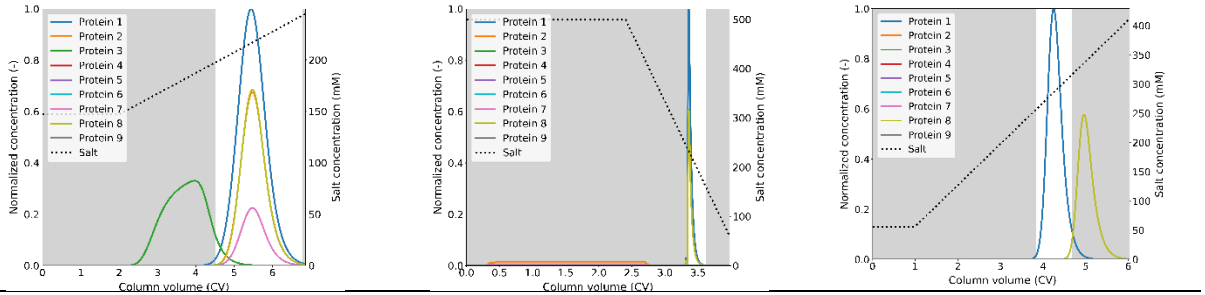
Simultaneous — **CEX – Dilution – HIC – Dilution – AEX**

Purity: 97.2%
Yield: 99.7%
Buffer cons.: 7.43 L/g
WOP: 97



Top-to-bottom — **AEX – Dilution – HIC – Dilution – CEX**

Purity: 99.3%
Yield: 95.7%
Buffer cons.: 5.35 L/g
WOP: 97



Decomposition — **AEX – Dilution – HIC – Dilution – CEX**

Purity: 99.2%
Yield: 96.0%
Buffer cons.: 5.42 L/g
WOP: 97

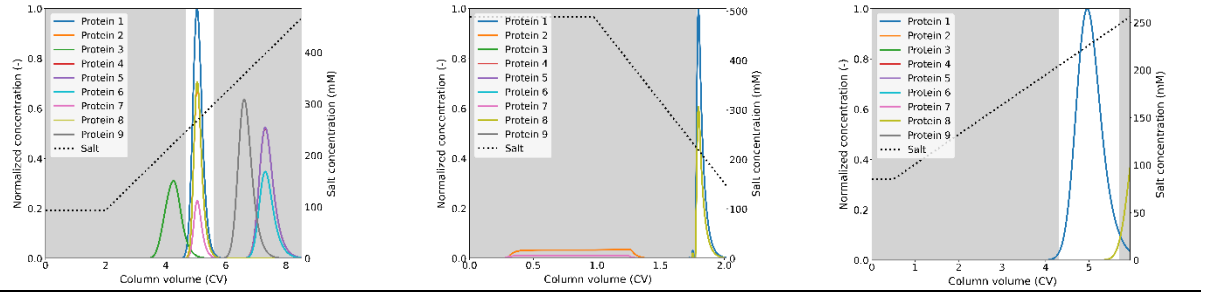


Figure 4. 7. Final local optimization results using the simultaneous strategy with MMs. The global results of the best flowsheets for each strategy using MMs are used as input for the final local optimization. The maximum number of iterations was 50.

For comparing the overall computational effort, the total amount of hours for each strategy and workflow (MMs or ANNs) are evaluated and shown in Figure 4.8. However, the overall flowsheet optimization workflow applied parallelization whenever possible. The ANN-time involves the data-generation (using MMs), ANN development, and running the optimization, though, 99% of the time is devoted to the data-generation. The MM only includes the optimization time. The simultaneous strategy with MMs is obviously the most computationally intensive, whereas the top-to-bottom with MMs requires the least amount of computational effort.

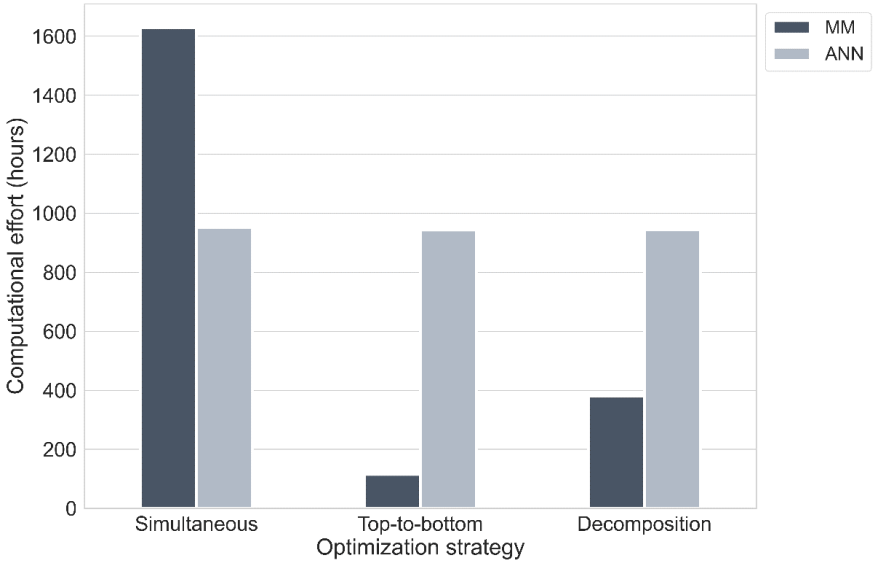


Figure 4. 8. Comparison of the overall computational effort between the optimization strategies and modeling workflows. The computational hours represent the total (sequential) amount of hours needed for each strategy.

Nowadays, more advanced computers consist of at least 10 or even 20 cores, and as a consequence the simultaneous and decomposition strategy can be executed way more time-efficiently. The decomposition can be parallelized maximally 15 times, as sequences of three chromatography steps depend on the two-chromatography step sequences. Whereas, the simultaneous strategy can be split into the number of flowsheets to be evaluated, in this case 39. Similarly for the ANN workflow, where, in principle, infinite codes can run simultaneously to generate data. Only the top-to-bottom strategy with MMs cannot be parallelized, as decisions are made sequentially between the various levels of chromatography steps. Figure 4.9 shows the effect of using 10 or 20 cores on each strategy and workflow. The decomposition strategy with MMs is the most time-efficient when making optimal use of the cores. In this case study, ANNs are significantly more time-efficient for the simultaneous strategy and for the top-to-bottom strategy when using 20 cores.

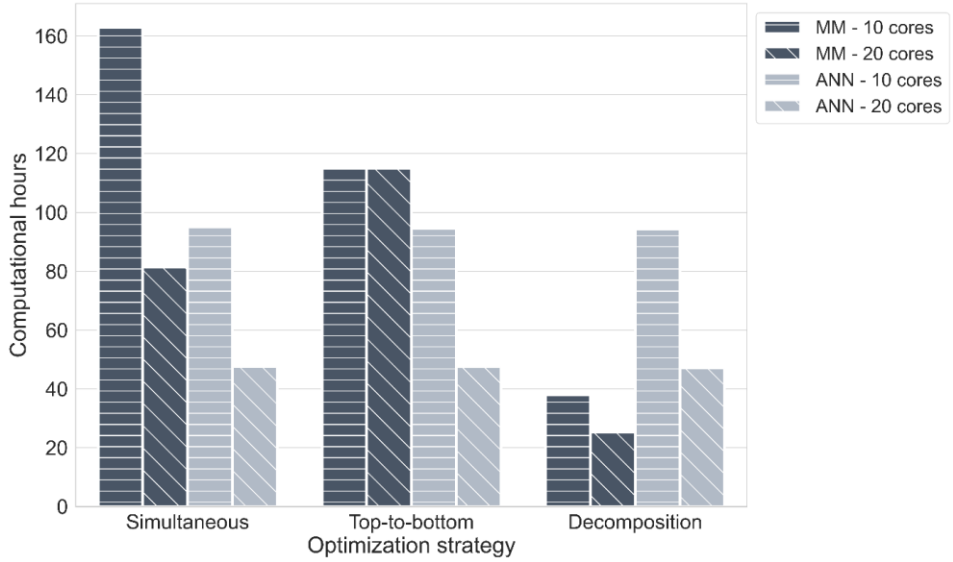


Figure 4. 9. Comparing the computational hours for each optimization strategy and modeling workflow when using 10 or 20 cores.

Evidentially, the optimization strategy plays a significant role in the overall computational effort. But, if the optimization strategy and workflow are parallelized most efficiently, the difference in computational time between the strategies decreases, ranging from about 1 to 7 days. In this case study, all strategies found multiple and similar optimal flowsheets. However, to obtain the most optimal conditions when connecting several unit operations, the simultaneous strategy is still recommended. In this flowsheet optimization evaluation, ANNs did not appeared to be more time-efficient. Presumably, if more resins and/or larger sequences are considered and at least 20 cores can be used, it is expected that the ANNs exceeds the time-efficiency compared to MMs. This would be an interesting evaluation for a follow-up. Moreover, ANNs are very fast in executing the flowsheet optimization, which can be advantageous when evaluating different scenarios for the optimization problem. In general, multiple factors determine which optimization strategy and workflow (MMs or ANNs) might be optimal for a specific case study, such as, the objective(s) and constraint(s), the size of the superstructure, and/or the computer power. The overview in Table 4.5 can help to make decisions for a flowsheet optimization approach.

Table 4. 5. Suggestions for deciding the type of optimization strategy and/or modeling workflow (ANNs or MMs) for certain scenarios/case studies.

Optimization problem

Optimization objectives and constraints

- Objective(s) and constraint(s) are clear:
  MMs, however, depending on superstructure size
- Different objective(s) and constraint(s) to be evaluated:
  ANNs

Superstructure size

Number of chromatography modes (type of resins) to be considered:

- 3 chromatography modes: MMs
- 4 chromatography modes: ANNs + MMs
- 5 chromatography modes: ANNs + MMs

Time

Depending on available number of cores.

If multiple cores can be used:

- Limited time:
  Decomposition strategy
- Extended time:
  Simultaneous strategy

Flexibility of method

Optimizing variables

- Decided variables:
  MMs and/or ANNs
- Undecided variables:
  MMs easier to use,
  or make more general ANNs with various input variables, or generate multiple ANNs

Apply different objectives for different steps

Decomposition strategy, this strategy can apply different objectives for the first step (capture step) and second and/or third steps (polishing steps).

Complexity

In terms of coding and knowledge

- All optimization strategies are about equal in development complexity, as the general optimization workflow is similar to all of them for both ANNs and MMs
- Developing the ANNs adds more complexity to the overall approach
- Advanced knowledge is required on the various MMs employed, the overall optimization workflow, developing the ANNs, all the algorithms/solvers used for the optimization and ANNs

In terms of solving

- Least complex:
  Top-to-bottom, as it individually solves each unit operation
- Most complex:
  Simultaneous, challenging to find the optimal solution for a sequence of more than 3 unit operations having at least 5 variables per unit operation. Increasing the number of unit operations in the sequence or the number of variables will significantly increase the complexity to solve the problem

4.4. Conclusions

In this study, we compared three optimization strategies to determine the most effective approach for complex flowsheet optimization based on their outcomes, time-efficiency, and complexity. Each strategy, e.g., simultaneous, top-to-bottom, and decomposition of the superstructure, was evaluated by either using MMs or ANNs for the global optimization. This complex flowsheet optimization consisted of 39 flowsheets, including an optional buffer exchange between the chromatography steps. The filtration mathematical model was validated for an UF/DF step using BSA. The protein concentration achieved an R^2 of 0.99 and a standard deviation of 0.03, and the salt concentration achieved an R^2 of 0.97 and a standard deviation of 6.25. Therefore, this model was assumed to be valid and applicable to the other proteins during flowsheet optimization, which had a similar or higher molecular weight than BSA. For the ANNs, all critical proteins, which are present around the product peak, reached an $R^2 > 0.93$, and the product of interest achieved an $R^2 > 0.98$ and RMSE < 0.022.

Subsequently, flowsheet optimization using MMs identified the same optimal flowsheet (AEX – D – HIC – D – CEX) for both top-to-bottom and decomposition strategies, the ANNs predicted the same WOP for this sequence. The simultaneous strategy with MMs identified a different sequence (CEX – D – HIC – D – AEX), which was not selected as one of the best by the other two strategies, giving a WOP threshold of at least 96. In general, the WOP values were predicted within a similar range when using either ANNs or MMs. In the case of orthogonal sequences, the solver often determined the optimal salt conditions to exclude the filtration step and instead employed a dilution / spiking step, and so reducing buffer consumptions and enhancing yields. Leveraging the multi-core processing capabilities, commonly available in contemporary computers, minimizes the duration of the flowsheet optimization between the strategies. When using multiple cores, the superstructure decomposition method employed with MMs is the most time-efficient approach. Utilizing ANNs is only significantly more time-efficient when employing the simultaneous strategy, and top-to-bottom approach when utilizing 20 cores. Furthermore, if various optimization problems want to be evaluated, ANNs are valuable for their fast flowsheet optimization, taking under an hour with multiple cores. All strategies are about equal in terms of complexity to develop the software. However, the combination with ANNs adds a layer of complexity because more knowledge is required on different aspects.

This study points out the importance of different optimization strategies and modeling techniques for complex flowsheet optimizations. Since numerous factors play a role, the decision-making table can support to find the most suitable type of strategy and modeling technique for a certain case study. Flowsheet optimization is crucial during the early

conceptual process design to decrease costs and development time. Moreover, at the initial stage of a development process, limited sample material is available and knowledge about the sample purification has yet to be acquired. All strategies, whether employing MMs and ANNs, successfully identified multiple optimal flowsheets. Moreover, due to efficient parallelization, the difference in computational time between the strategies was minimized. Though, the decomposition of the superstructure strategy with MMs proved to be most time-efficient. Furthermore, it has the advantage to apply different objectives for specific steps during the purification process, enhancing its versatility and utility in biopharmaceutical process development.

Acknowledgement

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. Moreover, the authors want to thank Dr. Ir. Tim Nijssen for the fruitful discussions on the optimization strategies and specifically on the decomposition strategy. The authors also want to thank Roxana Disela for performing the additional HIC chromatography experiments.

4.5. References

[1] G. Jagschies, K.M. Łącki, Chapter 4 - Process Capability Requirements, in: G. Jagschies, E. Lindskog, K. Łącki, P. Galliher (Eds.), Biopharmaceutical Processing, Elsevier2018, pp. 73-94. https://doi.org/https://doi.org/10.1016/B978-0-08-100623-8.00004-9.

[2] K.M. Łącki, Chapter 16 - Introduction to Preparative Protein Chromatography, in: G. Jagschies, E. Lindskog, K. Łącki, P. Galliher (Eds.), Biopharmaceutical Processing, Elsevier2018, pp. 319-366. https://doi.org/https://doi.org/10.1016/B978-0-08-100623-8.00016-5.

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

[4] T.C. Huuk, T. Hahn, A. Osberghaus, J. Hubbuch, Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography, Sep Purif Technol 136 (2014) 207-222. https://doi.org/https://doi.org/10.1016/j.seppur.2014.09.012.

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

[6] L. Mencarelli, Q. Chen, A. Pagot, I.E. Grossmann, A review on superstructure optimization approaches in process system engineering, Comput Chem Eng 136 (2020) 106808. https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.106808.

[7] Q. Chen, I.E. Grossmann, Recent Developments and Challenges in Optimization-Based Process Synthesis, Annual Review of Chemical and Biomolecular Engineering 8(1) (2017) 249-283. https://doi.org/https://doi.org/10.1146/annurev-chembioeng-080615-033546.

[8] J.J. Siirola, G.J. Powers, D.F. Rudd, Synthesis of system designs: III. Toward a process concept generator, AIChE Journal 17(3) (1971) 677-682. https://doi.org/https://doi.org/10.1002/aic.690170334.

[9] T. Umeda, A. Hirai, A. Ichikawa, Synthesis of optimal processing system by an integrated approach, Chem Eng Sci 27(4) (1972) 795-804. https://doi.org/https://doi.org/10.1016/0009-2509(72)85013-9.

[10] A.W. Westerberg, A retrospective on design and process synthesis, Comput Chem Eng 28(4) (2004) 447-458. https://doi.org/https://doi.org/10.1016/j.compchemeng.2003.09.029.

[11] L.K.E. Achenie, L.T. Biegler, A superstructure based approach to chemical reactor network synthesis, Comput Chem Eng 14(1) (1990) 23-40. https://doi.org/https://doi.org/10.1016/0098-1354(90)87003-8.

[12] M.H. Bauer, J. Stichlmair, Superstructures for the mixed integer optimization of nonideal and azeotropic distillation processes, Comput Chem Eng 20 (1996) S25-S30. https://doi.org/https://doi.org/10.1016/0098-1354(96)00015-4.

[13] M. Short, A.J. Isafiade, D.M. Fraser, Z. Kravanja, Synthesis of heat exchanger networks using mathematical programming and heuristics in a two-step optimisation procedure with detailed exchanger design, Chem Eng Sci 144 (2016) 372-385. https://doi.org/https://doi.org/10.1016/j.ces.2016.01.045.

[14] F. Friedler, K.B. Aviso, B. Bertok, D.C.Y. Foo, R.R. Tan, Prospects and challenges for chemical process synthesis with P-graph, Curr Opin Chem Eng 26 (2019) 58-64. https://doi.org/https://doi.org/10.1016/j.coche.2019.08.007.

[15] Q. Chen, Y. Liu, G. Seastream, J.D. Siirola, I.E. Grossmann, Pyosyn: A new framework for conceptual design modeling and optimization, Comput Chem Eng 153 (2021) 107414. https://doi.org/https://doi.org/10.1016/j.compchemeng.2021.107414.

[16] M.-O. Bertran, R. Frauzem, L. Zhang, R. Gani, A Generic Methodology for Superstructure Optimization of Different Processing Networks, in: Z. Kravanja, M. Bogataj (Eds.), Computer Aided Chemical Engineering, Elsevier2016, pp. 685-690. https://doi.org/https://doi.org/10.1016/B978-0-444-63428-3.50119-3.

[17] R. Altissimi, A. Brambilla, A. Deidda, D. Semino, Optimal operation of a separation plant using artificial neural networks, Comput Chem Eng 22 (1998) S939-S942. https://doi.org/https://doi.org/10.1016/S0098-1354(98)00185-9.

[18] M. Chambers, C.A. Mount-Campbell, Process optimization via neural network metamodeling, International Journal of Production Economics 79(2) (2002) 93-100. https://doi.org/https://doi.org/10.1016/S0925-5273(00)00188-2.

[19] C.A. Henao, C.T. Maravelias, Surrogate-based superstructure optimization framework, AIChE Journal 57(5) (2011) 1216-1232. https://doi.org/https://doi.org/10.1002/aic.12341.

[20] F.A.N. Fernandes, Optimization of Fischer-Tropsch Synthesis Using Neural Networks, Chem Eng Technol 29(4) (2006) 449-453. https://doi.org/https://doi.org/10.1002/ceat.200500310.

[21] A.M. Schweidtmann, A. Mitsos, Deterministic Global Optimization with Artificial Neural Networks Embedded, Journal of Optimization Theory and Applications 180(3) (2019) 925-948. https://doi.org/https://doi.org/10.1007/s10957-018-1396-0.

[22] C.A.O. Nascimento, R. Giudici, R. Guardani, Neural network based approach for optimization of industrial chemical processes, Comput Chem Eng 24(9) (2000) 2303-2314. https://doi.org/https://doi.org/10.1016/S0098-1354(00)00587-1.

[23] I.C. Reinhardt, D.J.C. Oliveira, D.D.T. Ring, Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector, Journal of Industrial Information Integration 18 (2020) 100131. https://doi.org/https://doi.org/10.1016/j.jii.2020.100131.

[24] F. Silva, D. Resende, M. Amorim, M. Borges, 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) (2020) 175. https://doi.org/https://doi.org/10.3390/joitmc6040175.

[25] M. Bisschops, L. Cameron, Process Intensification and Industry 4.0: Mutually Enabling Trends, Process Control, Intensification, and Digitalisation in Continuous Biomanufacturing2022, pp. 209-229. https://doi.org/https://doi.org/10.1002/9783527827343.ch7.

[26] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review, Processes 8(9) (2020) 1088. https://doi.org/https://doi.org/10.3390/pr8091088.

[27] R.M.C. Portela, C. Varsakelis, A. Richelle, N. Giannelos, J. Pence, S. Dessoy, M. von Stosch, When Is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to the Digital Twin Concept, Digital Twins, Springer Berlin Heidelberg, Berlin, Heidelberg, 2020, pp. 35-55. https://doi.org/https://doi.org/10.1007/10_2020_138.

[28] FDA, PAT Guidance for Industry - A Framework for innovative Pharmaceutical Development, Manufacturing and Quality Assurance, 2004. www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance.

[29] L.X. Yu, Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control, Pharmaceutical Research 25(4) (2008) 781-791. https://doi.org/https://doi.org/10.1007/s11095-007-9511-1.

[30] A. Felinger, G. Guiochon, Comparison of the Kinetic Models of Linear Chromatography, Chromatographia 60(1) (2004) S175-S180. https://doi.org/https://doi.org/10.1365/s10337-004-0288-7.

[31] V. Kumar, A.M. Lenhoff, Mechanistic Modeling of Preparative Column Chromatography for Biotherapeutics, Annual Review of Chemical and Biomolecular Engineering 11(1) (2020) 235-255. https://doi.org/https://doi.org/10.1146/annurev-chembioeng-102419-125430.

[32] L.K. Shekhawat, A. Tiwari, S. Yamamoto, A.S. Rathore, An accelerated approach for mechanistic model based prediction of linear gradient elution ion-exchange chromatography of proteins, Journal of Chromatography A 1680 (2022) 463423. https://doi.org/https://doi.org/10.1016/j.chroma.2022.463423.

[33] D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch, Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications, Biotechnology Progress n/a(n/a) (2020) e2984. https://doi.org/https://doi.org/10.1002/btpr.2984.

[34] R. Hess, D. Yun, D. Saleh, T. Briskot, J.-H. Grosch, G. Wang, T. Schwab, J. Hubbuch, Standardized method for mechanistic modeling of multimodal anion exchange chromatography in flow through operation, Journal of Chromatography A 1690 (2023) 463789. https://doi.org/https://doi.org/10.1016/j.chroma.2023.463789.

[35] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/https://doi.org/10.1002/bit.24576.

[36] E.J. Close, J.R. Salm, D.G. Bracewell, E. Sorensen, A model based approach for identifying robust operating conditions for industrial chromatography with process variability, Chem Eng Sci 116 (2014) 284-295. https://doi.org/https://doi.org/10.1016/j.ces.2014.03.010.

[37] D. Gétaz, G. Stroehlein, A. Butté, M. Morbidelli, Model-based design of peptide chromatographic purification processes, Journal of Chromatography A 1284 (2013) 69-79. https://doi.org/https://doi.org/10.1016/j.chroma.2013.01.118.

[38] R. Disela, O.L. Bussy, G. Geldhof, M. Pabst, M. Ottens, Characterisation of the E. coli HMS174 and BLR host cell proteome to guide purification process development, Biotechnology Journal 18(9) (2023) 2300068. https://doi.org/https://doi.org/10.1002/biot.202300068.

[39] K. Meyer, S. Leweke, E. von Lieres, J.K. Huusom, J. Abildskov, ChromaTech: A discontinuous Galerkin spectral element simulator for preparative liquid chromatography, Comput Chem Eng 141 (2020) 107012. https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.107012.

[40] S. Leweke, E. von Lieres, Chromatography Analysis and Design Toolkit (CADET), Comput Chem Eng 113 (2018) 274-294. https://doi.org/https://doi.org/10.1016/j.compchemeng.2018.02.025.

[41] J.M. Breuer, S. Leweke, J. Schmölder, G. Gassner, E. von Lieres, Spatial discontinuous Galerkin spectral element method for a family of chromatography models in CADET, Comput Chem Eng 177 (2023) 108340. https://doi.org/https://doi.org/10.1016/j.compchemeng.2023.108340.

[42] J.S. Rao, A. Püttmann, S. Khirevich, U. Tallarek, C. Geuzaine, M. Behr, E. von Lieres, High-definition simulation of packed-bed liquid chromatography, Comput Chem Eng 178 (2023) 108355. https://doi.org/https://doi.org/10.1016/j.compchemeng.2023.108355.

[43] H. Narayanan, M. von Stosch, F. Feidl, M. Sokolov, M. Morbidelli, A. Butté, Hybrid modeling for biopharmaceutical processes: advantages, opportunities, and implementation, Frontiers in Chemical Engineering 5 (2023). https://doi.org/https://doi.org/10.3389/fceng.2023.1157889.

[44] M. von Stosch, R. Oliveira, J. Peres, S.F. de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput Chem Eng 60 (2014) 86-101. https://doi.org/https://doi.org/10.1016/j.compchemeng.2013.08.008.

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

[46] A.A. Kiss, J. Grievink, Process systems engineering developments in Europe from an industrial and academic perspective, Comput Chem Eng 138 (2020) 106823. https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.106823.

[47] S. Liu, L.G. Papageorgiou, Optimal Antibody Purification Strategies Using Data-Driven Models, Engineering 5(6) (2019) 1077-1092. https://doi.org/https://doi.org/10.1016/j.eng.2019.10.011.

[48] J.M. Natali, J.M. Pinto, L.G. Papageorgiou, Efficient MILP formulations for the simultaneous optimal peptide tag design and downstream processing synthesis, AIChE Journal 55(9) (2009) 2303-2317. https://doi.org/https://doi.org/10.1002/aic.11913.

[49] E.M. Polykarpou, P.A. Dalby, L.G. Papageorgiou, Optimal synthesis of chromatographic trains for downstream protein processing, Biotechnology Progress 27(6) (2011) 1653-1660. https://doi.org/https://doi.org/10.1002/btpr.670.

[50] S. Liu, L.G. Papageorgiou, Multi-objective optimisation for biopharmaceutical manufacturing under uncertainty, Comput Chem Eng 119 (2018) 383-393. https://doi.org/https://doi.org/10.1016/j.compchemeng.2018.09.015.

[51] E. Simeonidis, J.M. Pinto, M.E. Lienqueo, S. Tsoka, L.G. Papageorgiou, MINLP Models for the Synthesis of Optimal Peptide Tags and Downstream Protein Processing, Biotechnology Progress 21(3) (2005) 875-884. https://doi.org/https://doi.org/10.1021/bp049650n.

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

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

[54] D. Keulen, E. van der Hagen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Using artificial neural networks to accelerate flowsheet optimization for downstream process development, Biotechnology and Bioengineering (2023) 1-14. https://doi.org/https://doi.org/10.1002/bit.28454.

[55] P. Tanartkit, L.T. Biegler, A nested, simultaneous approach for dynamic optimization problems—I, Comput Chem Eng 20(6) (1996) 735-741. https://doi.org/https://doi.org/10.1016/0098-1354(95)00206-5.

[56] SciPy, scipy.optimize.differential_evolution - SciPy v1.11.2 Reference Guide, 2023. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html.

[57] G.R. Kocis, I.E. Grossmann, A modelling and decomposition strategy for the minlp optimization of process flowsheets, Comput Chem Eng 13(7) (1989) 797-819. https://doi.org/https://doi.org/10.1016/0098-1354(89)85053-7.

[58] M.M. Daichendt, I.E. Grossmann, Integration of hierarchical decomposition and mathematical programming for the synthesis of process flowsheets, Comput Chem Eng 22(1) (1998) 147-175. https://doi.org/https://doi.org/10.1016/S0098-1354(97)88451-7.

[59] D.A. Liñán, D.E. Bernal, L.A. Ricardez-Sandoval, J.M. Gómez, Optimal design of superstructures for placing units and streams with multiple and ordered available locations. Part I: A new mathematical framework, Comput Chem Eng 137 (2020) 106794. https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.106794.

[60] D.M. Ruthven, Principles of adsorption and adsorption processes, John Wiley & Sons, New York, 1984.

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

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

[63] M. Fellner, A. Delgado, T. Becker, Functional nodes in dynamic neural networks for bioprocess modelling, Bioproc Biosyst Eng 25(5) (2003) 263-270. https://doi.org/https://doi.org/10.1007/s00449-002-0297-6.

[64] G.A. Foley, Membrane Filtration: A Problem Solving Approach with MATLAB, 2013.

[65] L. Petzold, Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations, SIAM Journal on Scientific and Statistical Computing 4(1) (1983) 136-148. https://doi.org/https://doi.org/10.1137/0904010.

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

Chapter 5

From protein structure to an optimized chromatographic capture step using multiscale modeling

Optimizing a biopharmaceutical chromatographic purification process is currently the greatest challenge during process development. A lack of process understanding calls for extensive experimental efforts in pursuit of an optimal process. *In silico* techniques, such as mechanistic or data driven modeling, enhance the understanding, allowing more cost-effective and time efficient process optimization. This work presents a modeling strategy integrating quantitative structure property relationship (QSPR) models and chromatographic mechanistic models (MM) to optimize a cation exchange (CEX) capture step limiting experiments. In QSPR, structural characteristics obtained from the protein structure are used to describe physicochemical behavior. This QSPR information can be applied in MM to predict the chromatogram and optimize the entire process. To validate this approach, retention profiles of six proteins were determined experimentally from two mixtures, at different pH (3.5, 4.3, 5.0, 7.0). Four proteins at different pH's were used to train QSPR models predicting the retention times and characteristic charge, subsequently the equilibrium constant was determined. For an unseen protein knowing only the protein structure, the retention peak difference between the modeled and experimental peaks only was 0.2% relative to the gradient length (60 column volume). Subsequently, the CEX capture step was optimized, demonstrating a consistent result in both the experimental and QSPR-based methods. The impact of model parameter confidence on the final optimization revealed two viable process conditions, one of which is similar to the optimization achieved using experimentally obtained parameters. The multiscale modeling approach reduces the required experimental effort by identification of initial process conditions which can be optimized.

This chapter has been submitted for publication: Keulen, D., Neijenhuis, T., Lazopoulou, A., Disela, R., Geldhof, G., Le Bussy, O., Klijn, M.E., and Ottens, M..

5.1. Introduction

Over the past years, the biopharmaceutical industry has experienced substantial growth, with protein-based biopharmaceuticals (e.g., monoclonal antibodies (mAbs) and protein subunit vaccines) being a significant part of the industry [1]. As a consequence, the biopharmaceutical industry endeavors to accelerate process development with the primary goal to deliver biopharmaceuticals at the earliest possible time, pushing the competitive market [2]. Moreover, the competition even intensified more due to the emerging field of biosimilars [3, 4]. The biopharmaceutical sector requires therefore innovative approaches to advance process development, while ensuring product quality and stability. Especially the downstream process is the major cost driver of the overall manufacturing costs, demanding an efficient and cost-effective process. To achieve very high product purities, chromatography is currently the most essential but also the most costly technique [5].

In silico techniques, such as mechanistic or data-driven modeling, can be of great merit for process development. These methods allow for increased process understanding while reducing experimental effort and/or use of critical sample material, and decreasing process development times [6, 7]. Within the next years, modeling techniques will become more essential for biopharmaceutical industry. Specifically for Industry 4.0 that aims to digitalize the entire manufacturing process [8-11]. Moreover, increased process understanding and process and product quality control are in agreement with the Quality-by-Design (QbD) guidelines [12-15]. Identifying the operating window of the critical process parameters (CPP) is an essential part to guarantee process' stability. Currently, these operating windows are determined with expensive and time-consuming wet-lab Design-of-Experiments (DoE). Chromatographic MM attempt to describe the chromatographic process *in silico* and could be an inexpensive and fast alternative to determine the CPP operating window. Over the past years, the industry has been gradually adopting chromatographic MM, with ongoing advancement being made in determining the essential input parameters [16-18]. In the future, the ultimate objective is to determine adsorption isotherm for complex mixtures more easily [19, 20]. Progress in utilizing mass spectrometry data could play a crucial role in achieving this goal [21]. However, at this moment determining adsorption isotherm parameters for the MM remains a bottleneck for industrial application, mainly due to time and material limitations especially in the early phase of downstream process development [22]. Quantitative Structure Property Relationships (QSPR) modeling could be an in silico alternative to experimentally determining the adsorption isotherm parameters. QSPR aims to correlate physicochemical properties with specific behavior, such as chromatographic retention time [23]. These physicochemical properties are calculated from protein structure models that describe the position of each

atom. Combining MM with QSPR and optimization tools could pave the way for a holistic modeling approach/workflow.

In 2001, Mazza et al. introduced a QSPR model for predicting protein retention times for ion exchange chromatography [23]. Their approach involved feature calculation using the proprietary software platform MOE a genetic algorithm for feature selection for the training of a partial least squares model [24, 25]. As a result, several follow-up studies applied QSPR models to different modes of chromatography/type of chromatography resins, using support vector machine regression methods, and including pH effects [26-31]. Malmquist et al. developed an additional set of protein descriptors that are pH-dependent and based on electrostatic and hydrophobic properties [32]. Moreover, several studies considered the crucial binding orientations within protein-resin binding affinities in their QSPR models [33-35]. In recent years, QSPR has been applied to more complex proteins, such as Fabs and mAbs, showing the growing interest from industry and the added value of these models [22, 36, 37]. Robinson et al. showed the potential of QSPR models for *in silico* resin screening of six chromatographic systems applied to Fabs [36]. While Saleh et al. built QSPR models using 21 mAbs variants to predict the adsorption isotherm parameters, the equilibrium constant and the characteristic charge, which were subsequently applied to the MM and able to predict the cation exchange chromatography (CEX) step [22]. Their study shows promising capabilities of a multiscale model to simulate different process conditions without the need for wet-lab experiments. Several software packages are available to calculate the protein descriptors that are needed for QSPR modeling, an overview of these software packages has been provided elsewhere [38, 39]. Most software tools are only available via webservers or commercially, lacking source code availability. Therefore, Neijenhuis et al. have recently published an open-source QSPR software tool, which has also been used in this work [40].

Most research on QSPR modeling either developed protein descriptors or applied existing protein descriptors for their QSPR model with the aim to increase the protein-behavior understanding via retention prediction [29, 32, 36, 37, 41]. Additionally, other research also applied the predicted QSPR parameters to MM and validated the predicted chromatographic process from a protein structure/sequence [22, 28, 30]. So far, no research has shown the ability of QSPR models in combination with MM to optimize a chromatographic process step without any need for protein material. Moreover, the influence of the accuracy of the predicted QSPR-parameters on an optimized process has not yet been evaluated.

This chapter presents a general multiscale modeling strategy that integrates QSPR and chromatographic MM to optimize a CEX capture step. We were able to simulate and validate a CEX step only using the protein structure. Subsequently, we compared the uncertainty of

the experimentally determined and predicted parameters on the final optimization outcome. An overview of the experimental-based and QSPR-based strategy is shown in Figure 5.1. This strategy can be used to determine the operating window of CPPs in an early stage process development, showing the potential applicability for industry. Combining these modeling techniques together with an optimization software reduces the experimentally effort overall process development time significantly. Previous research mostly used pure components to perform the linear gradient experiments (LGE), however the availability of pure components is limited in biopharmaceutical industry. Therefore, performing LGE with complex protein mixtures would offer significant advantages. So far, only Buyel et al. applied QSPR modeling to a crude mixture of plant extracts to predict retention times for ion exchange and mixed mode chromatography separations [31]. Here, we performed LGE for five different gradient lengths and four pHs applied to two mixtures of each three proteins. Performing the experiments with protein mixtures instead of each protein individually, reduces the total LGE from 30 to 10 experiments. We developed QSPR models for predicting the retention times and characteristic charges. These predicted QSPR parameters were used to obtain the equilibrium constants. The multiscale model was validated for an unseen protein, which was excluded from the QSPR training and testing data. Finally, we compared the influence of parameter uncertainties on the optimization outcome by using experimental and QSPR predicted parameters.

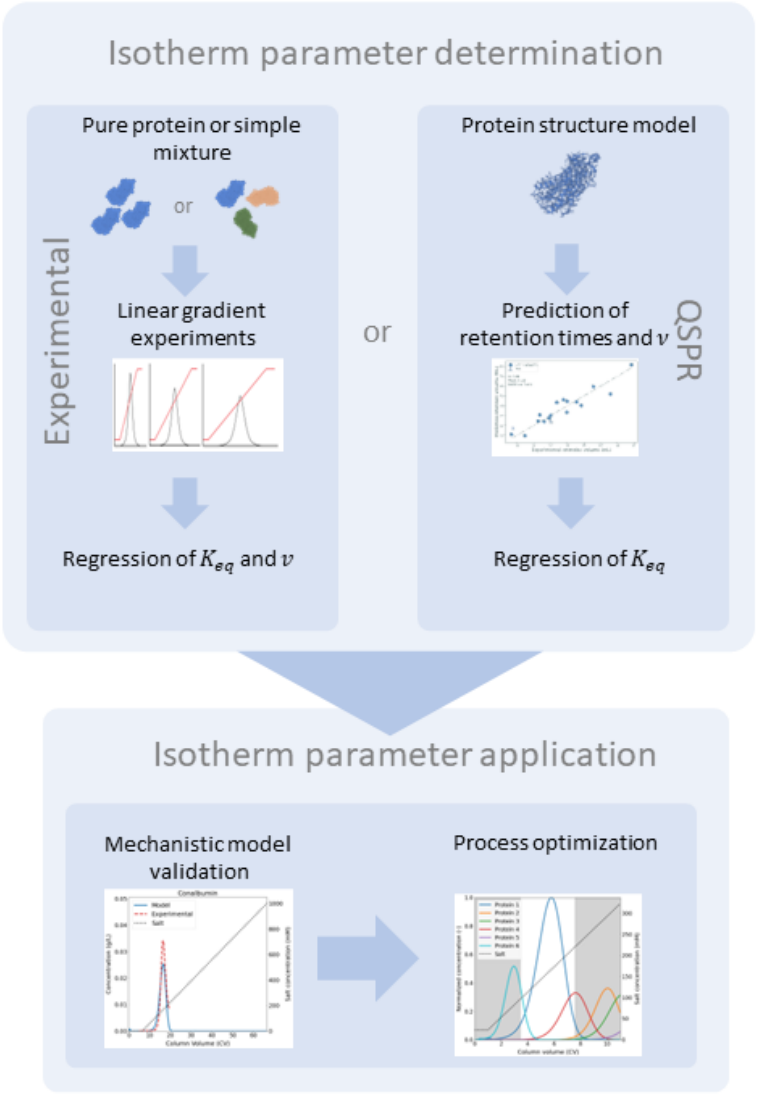


Figure 5. 1. Overview of the experimental-based method and the QSPR-based method. Both methods can be used to determine the adsorption isotherm parameters that can be used in the mechanistic model for process optimization purposes. The equilibrium constant is denoted by K_{eq} and the stoichiometric coefficient of salt counter ions with ν.

5.2. Materials & Methods

5.2.1. Experimental part

5.2.1.1. Materials & Equipment

A 1-mL CEX column of HiTrap SP FF (Cytiva Life Sciences, USA) was used for the preparative column experiments. For the analytical size exclusion chromatography – ultra performance liquid chromatography (SEC-UPLC), an ACQUITY UPLC Protein BEH SEC 200 Å column (Water Corporation, USA) was used, protected with a prior/foregoing ACQUITY UPLC Protein BEH SEC guard 200 Å column (Water Corporation, USA).

The following proteins were purchased from Sigma-Aldrich, USA: bovine serum albumin (BSA), lysozyme, cytochrome C, chymotrypsinogen A from bovine pancreas, and conalbumin. Ribonuclease pancreatic (RNase) was purchased from Roche Diagnostics GmbH, Germany. Dextran (DXT1740K) (American Polymer Standards Corporation, USA) was used for column characterization.

The buffers were prepared with Milli-Q water and adjusted to the desired pH using either 0.5 M sodium hydroxide or 1 M hydrochloric acid. The buffers were filtered to remove undissolved salts, 0.2 μm pore-size hollow fiber MediaKap (Repligen, USA) filter for UPLC buffers and a 0.2 μm Membrane Disc Filter (Pall corporation, USA) for ÄKTA buffers. Moreover, all buffers were degassed for 20 minutes using an ultrasonic bath (Branson Ultrasonics, USA) to prevent introducing air bubbles into the column. The protein mixture was filtered using a 0.2 μm Whatman Puradisc FP 30 mm (GE Healthcare Life Sciences, USA).

5.2.1.2. Linear gradient column experiments

LGE were conducted at various pH values (pH 3.5, 4.3, 5.0, and 7.0) for five gradient lengths: 20, 30, 40, 60, and 80 column volumes (CV). For every pH a different running buffer was needed, citric acid monohydrate (pH 3.5, 20 mM), sodium acetate trihydrate (pH 4.3 and 5.0, 50 mM), and sodium phosphate monobasic dihydrate (pH 7.0, 50 mM). The elution buffer is the same as the running buffer for that respective pH with the addition of 1 M sodium chloride. The pH-values were selected to theoretically favor a positive net charge for most proteins, and therefore anticipating their binding to the CEX resin. The chromatographic column experiments were performed on an ÄKTA pure system (Cytiva Life Sciences, USA) with UNICORN version 7.5 software, with a flowrate of 1 mL/min, and measuring UV absorbance at 230, 280, and 400 nm wavelength. The column characteristics are given in Table 5.1, more information on the characterization methods can be found in Appendix 5.A. During the chromatography runs, 1 mL samples were collected using a fraction collector. These samples were additionally analyzed with a Dionex UPLC system using Chromeleon Chromatography

Data System version 7 software, measuring UV absorbance at 230, 280, and 400 nm wavelength. The UPLC-running buffer was a 100 mM sodium phosphate monobasic dihydrate with a pH of 6.8. A flowrate of 0.1 mL/min and analysis time of 40 minutes was applied. The SEC-UPLC analysis enabled the identification of the peaks obtained during the LGE's with their corresponding proteins. However, the protein mixture was divided into two groups, as some proteins with similar characteristics were indistinguishable in the SEC-UPLC analysis. Group one consisted of RNase, cytochrome C, conalbumin, and group two of chymotrypsinogen, lysozyme, and albumin. Both multi-component mixtures contained 0.8 mg/mL of each protein.

Table 5. 1. Column characteristics for HiTrap SP FF column.

Parameter	Value	Unit
Column volume	0.97	mL
Column diameter[1]	0.70	cm
Bed height[1]	2.50	cm
Maximum pressure[1]	2.0	MPa
Ionic capacity[2]	800	mM
Particle size[1]	90	μm
Pore diameter[3]	54	nm
Cross sectional area	0.39	cm^2
System dead volume $(\boldsymbol{V_{dead}})$	0.34	mL
Total porosity $(\boldsymbol{\varepsilon_t})$	0.918	-
Extraparticle porosity $(\boldsymbol{\varepsilon_b})$	0.298	-
Intraparticle porosity $(\boldsymbol{\varepsilon_p})$	0.887	-
System dwell volume $(\boldsymbol{V_{dwell}})$	1.09	mL

[1]Manufacturer, [2]Osberghaus et al. [42], [3]Hagemann et al. [43].

Table 5. 2. Overview of the protein characteristics, in which PDB stands for Protein Data Bank.

Protein	PDB names	Mass (kDa)	Estimated Isoelectric point
Conalbumin	1OVT	75.83	6.62
Albumin	6QS9	66.43	5.49
Chymotrypsinogen	2CGA	25.67	8.13
Lysozyme	1GWD	14.31	9.20
RNase	1RNAse	13.69	8.29
Cytochrome C	6FF5	12.33	9.60

First, the column was equilibrated with 5 CV running buffer, followed by a 300 μL sample injection using a 10 mL Superloop (Cytiva Life Sciences, USA). After the sample injection, unretained proteins were removed by washing the column for 5 CV using the running buffer. Subsequently, a gradient elution was performed from 0 (running buffer) to 1 M sodium

chloride (elution buffer). The proteins in the collected fractions were identified with the SEC-UPLC analytical method. Though, it is expected that the elution order of the proteins remains the same and therefore, only the fractions of two gradients for each pH were analyzed with SEC-UPLC. For each fraction analysis, 5 μL sample was injected.

5.2.2. Chromatographic mechanistic model

The chromatographic MM from previous work was used to describe the dynamic adsorption behavior during the chromatographic separation process [44]. This employed MM is a combination of the equilibrium transport dispersive model combined with the linear driving force model as

$$\frac{\partial C_i}{\partial t} + F\frac{\partial q_i}{\partial t} = -u\frac{\partial C_i}{\partial x} + D_{L,i}\frac{\partial^2 C_i}{\partial x^2}, \qquad \textit{Eq. 5.1}$$

$$\frac{\partial q_i}{\partial t} = k_{ov,i}\left(C_i - C^*_{eq,i}\right), \qquad \textit{Eq. 5.2}$$

$$k_{ov,i} = \left[\frac{d_p}{6k_{f,i}} + \frac{d_p^2}{60\varepsilon_p D_{p,i}}\right]^{-1}, \qquad \textit{Eq. 5.3}$$

where the concentration in the liquid phase is represented by C_i and in the solid phase with q_i, in which subscript i denotes the protein component. The liquid phase concentration at equilibrium is denoted by $C^*_{eq,i}$. The phase ratio is equal to $F = (1-\varepsilon_b)/\varepsilon_b$, where ε_b is the bed porosity. Time and space are indicated by t and x respectively. u is the mobile phase interstitial velocity and D_L is the axial dispersion coefficient. The overall mass transfer coefficient, $k_{ov,i}$, is defined as the combined result of both the separate film mass transfer resistance and the mass transfer resistance within the pores [45]. In Eq. 5.3, the particle diameter is denoted by d_p, the intraparticle porosity by ε_p, and the effective pore diffusivity coefficient by D_p. The film mass transfer resistance is $k_f = D_f Sh/d_p$, in which D_f is the free diffusivity and Sh is the Sherwood number. The Livermore Solver for Ordinary Differentia Equations (LSODA) algorithm, part of the scipy.integrate package, is employed to solve the Ordinary Differential Equations (ODEs), automatically transitioning between the nonstiff Adams method and the stiff BDF method [46]. Additional details regarding the MM can be found in a prior study [47].

We employed the linear multicomponent mixed-mode isotherm, developed by Nfor et al. [48] to determine the equilibrium liquid phase concentration as

$$\frac{q_i}{C^*_{eq,i}} = K_{eq,i}\Lambda^{(v_i+n_i)}(z_s c_s)^{-v_i} c_v{}^{-n_i}\gamma_i, \qquad \textit{Eq. 5.4}$$

where the equilibrium constant, $K_{eq,i}$, quantifies the strength of the interaction between the protein and the stationary phase. Λ is the ligand density or ionic capacity of the concerned resin, z_s is the charge of the salt counter ion, c_s is the salt concentration in the liquid phase, and c_v is the molarity of the solution in the pore volume. The stoichiometric coefficient of salt counter ions is denoted by v_i, determined by $v_i = z_p/z_s$, in which z_p is the effective binding charge of the protein. For monovalent counter-ions, the charge equals one ($z_s = 1$), for example Na^+ in the sodium chloride elution buffer. In this work, only the ion-exchange part of the mixed-mode isotherm is used, therefore hydrophobic interaction stoichiometric coefficient (n_i) will be equal to zero. The activity coefficient (γ) of the protein solution can be calculated as

$$\gamma_i = e^{K_{s,i}c_s + K_{p,i}C_i}, \qquad \textit{Eq. 5.5}$$

where K_s is the salt-protein interaction constant and K_p the protein-protein interaction constant. In the linear range of adsorption, the protein concentrations are low and protein-protein interactions are expected to be minimal, therefore K_p becomes insignificant and can be neglected [49, 50]. Because of the low salting-out effects, the K_s also becomes negligible [49]. Subsequently, incorporating the assumptions for this work, the linear multicomponent mixed-mode isotherm is reformulated as

$$\frac{q_i}{C^*_{eq,i}} = K_{eq,i}\Lambda^{v_i}(z_s c_s)^{-v_i}\,. \qquad \textit{Eq. 5.6}$$

5

5.2.3. Procedure to determine adsorption isotherm parameters

The peak retention volumes were obtained from the LGE's for each gradient length and at each pH. The initial retention volumes ($V_{R,0}$) were corrected to be aligned with the elution gradients as follows:

$$V_R = V_{R,0} - V_m - V_D - \frac{V_{inj}}{2}, \qquad \textit{Eq. 5.7}$$

where V_R is the peak retention volum, V_m is the column void volume, determined by dextran pulse, and V_D is the system's dwell and dead volume, details can be found in Appendix 5.A [51]. The injection volume is denoted by V_{inj}, half of this volume needs to be subtracted [52].

The regression formula of Shukla et al. [53], adapted from Parente and Wetlaufer [51], was used to obtain the equilibrium constant (K_{eq}) and the characteristic charge (v) for each protein as follows:

$$V_R = \left(\left(C_{s,0}^{v+1} + \frac{V_m K_{eq} F \Lambda^v (v+1) * (C_{s,f} - C_{s,0})}{V_G}\right)^{\frac{1}{v+1}} - C_{s,0}\right) * \frac{V_G}{C_{s,f} - C_{s,0}}, \qquad \textit{Eq. 5.8}$$

where V_G is the gradient length. $C_{s,0}$ and $C_{s,f}$ are the initial and final salt concentration during the elution respectively. As no separate pore balance is considered in the chromatographic MM, the column phase ratio is considered the same $F = (1 - \varepsilon_b)/\varepsilon_b$. To validate the regression and accordingly the MM, the experimental data of 60 CV is left out during the regression.

The initial peak retention volumes ($V_{R,0}$) were determined using the function find_peaks of the signal module from the *Scipy* library. The regression was performed using the curve_fit function of the optimize module from the *Scipy* library.

Specifically at pH 5.0, Cytochrome C and RNAse co-eluted. The absorbance and respective calibration lines of Cytochrome C at 400 and 280 nm were used to trace back the RNAse peak. Moreover, at pH 4.3, Albumin and Chymotrypsinogen co-eluted. However, from the SEC-UPLC analysis it was observed that Albumin eluted later compared to the UV peak detected by the UNICORN software. Therefore, the peak retention volumes for Albumin at pH 4.3 were determined by analyzing the concentrations by SEC-UPLC in the 1 mL fractions obtained from the LGE. Albumin peak areas obtained from the SEC-UPLC were used to fit a third degree polynomial function representing the retention time as the maximum.

5.2.4. QSPR model

5.2.4.1. Structure preparation and descriptor calculation

For each protein, the respective models, listed in Table 5.2, were obtained from the protein data bank [54], specific entry selection was performed based on resolution and coverage. Duplicate chains were removed from each structural model using pdb-tools [55] to yield monomer representations. The side chain pKa of titratable residues were predicted using PROPKA3.0 [56] allowing for more accurate charge calculations with respect to pH. Protein features at pH 3.5, 4.3, 5.0 and 7.0 were calculated using our open-source software package prodes by Neijenhuis et al. [40] using the default settings, only supplying the modified pKa estimations. Visualization of protein structures was performed using UCSF-Chimera [57].

5.2.4.2. QSPR model training

For predicting the protein retention times and adsorption isotherm parameters, Multi Linear Regression (MLR) models were trained. The prediction of conalbumin was removed from the dataset prior to train-test splitting to eliminate all bias. To find an accurate predictive MLR model, series of filter thresholds were screened by testing a range of feature-feature correlation filters (Pearson correlations of 0.8, 0.9 and 0.99). Followed by feature-observation correlations filtering, maintaining a predefined percentage of features (10% to 100% in 10% increments). Feature selection was performed by sequential forward selection. Final models

were selected based on the cross-validated R^2 and test set RMSE, which should be close to the cross-validation RMSE to ensure model robustness. Feature importance was assessed by analysis of the regression coefficient and the influence of feature permutation. For the prediction of the unknown conalbumin, the confidence interval was calculated as

$$\hat{y}_h \pm t_{\left(1-\frac{\alpha}{2},n-p\right)} \times \sqrt{MSE\ (1 + X_h^T (X^T X)^{-1} X_h)}, \qquad \textit{Eq. 5.9}$$

where $\hat{y}_h$ is the predicted value, $t_{(1-\frac{\alpha}{2},n-p)}$ is the "t-multiplier", X and X_h are the feature matrixes of the training set and the value to be predicted. The mean squared error (MSE) is calculated as

$$MSE = \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2. \qquad \textit{Eq. 5.10}$$

5.2.5. Optimization

We evaluated the uncertainty-influence of the regressed and predicted QSPR adsorption isotherm parameters on the final optimization outcome. The equilibrium constant and characteristic charge values were varied between their standard deviation values for 100 samples. These samples were used in the optimization. First, the optimization was formulated and evaluated to be consistent when performing the same optimization multiple times. The global and local objectives were formulated as follows:

$$minf(x) = 2 * (100 - yield(x)) + 1 * \big(100 - purity(x)\big) \qquad \textit{Eq. 5.11}$$

$$s.t.\ \ h(x) = 0 \qquad \textit{Eq. 5.12}$$

$$0 \leq x \leq 1, \qquad \textit{Eq. 5.13}$$

where the objective function, $f(x)$, is minimized. The equality equations, such as the mass balances and equilibrium relations, need to be satisfied (Eq. 5.12). Moreover, variables (x) were normalized for more efficient optimization purposes (Eq. 5.13). Four variables were chosen namely, the initial and final salt concentrations, and the lower and upper cut points. The weights of the objective function were chosen to reflect a capture step to be optimized, hence removing most of the bulk impurities and preventing losing product material.

For the global optimization, the differential_evolution algorithm from the scipy.optimize package was employed, using the Latin hypercube sampling to initialize the population and the maximum number of iterations was 10 with a population size of 23. For the local optimization the Nelder-Mead algorithm was used, with a maximum of 100 iterations. The relative and function tolerances for both global and local optimizations were set to 1e-2. The lower cut point ranges from 1 – 80% on the left of the peak maximum, and the upper cut point

from 20 – 99% on the right of the peak maximum. The initial salt concentration varies between 1 – 150 mM, and the final salt concentration between 320 – 800 mM.

5.3. Results & Discussion

5.3.1. Linear gradient experiments

5.3.1.1. Determining the retention volume

LGE's were conducted for two protein mixtures at four pH values (pH 3.5, 4.3, 5.0, and 7.0) and various gradient lengths (20, 30, 40, 60, and 80 CV), as described in the experimental section 5.2.1. The elution order of the proteins was identified by SEC-UPLC analysis for each pH, to determine single peak retention volumes. The results for the 20 CV LGE are shown in Figure 5.2. As expected, a downward trend for the retention is observed when increasing the pH. No correlation between isoelectric point (PI) and retention was observed. Although cytochrome C, lysozyme, RNase and chymotrypsinogen elute in the order of descending pI (9.60, 9.20, 8.29 and 8.13 respectively) at pH 3.5. No retention volume for albumin and conalbumin (pI of 5.49 and 6.62, respectively) was determined as these proteins did not elute during the salt gradient, showing greater affinity for the column, which is in accordance with [30].

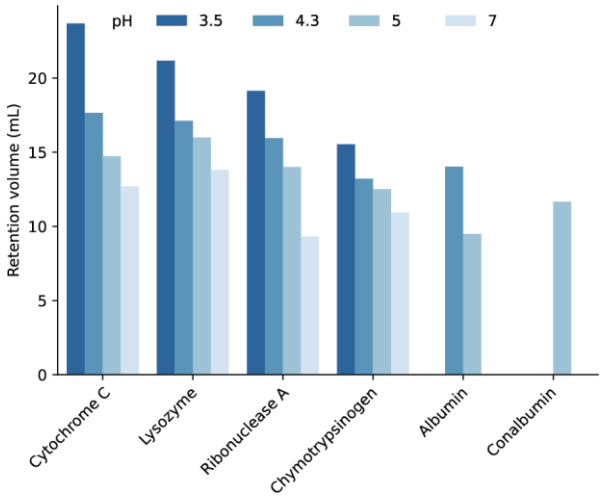


Figure 5. 2. Peak retention volumes (mL, y-axis) given for each protein (x-axis) at each pH (bars). These retention volumes are from the 20 CV gradient length using a HiTrap SP FF column, 1 CV is equal to 0.97 mL.

5.3.1.2. Regression of adsorption isotherm parameters

The corrected retention volumes, according to Eq. 5.7, were used to regress K_{eq} and ν using Eq. 5.8. The regression parameters for each protein at each pH are shown in Table 5.3. The regression plots of each protein at each pH are provided in Appendix 5.B, all fits achieved an R^2 close to one and RMSE values varied between 0.002 and 0.11.

Table 5. 3. Regressed adsorption isotherm parameters, the characteristic charge and the equilibrium constant, for each protein at each pH. The standard deviation is indicated with number after ± sign.

	Characteristic charge (ν)				Equilibrium constant (K_{eq})			
Protein	**pH 3.5**	**pH 4.3**	**pH 5**	**pH 7**	**pH 3.5**	**pH 4.3**	**pH 5**	**pH 7**
Conalbumin			2.37 ± 0.12				0.071 ± 0.02	
Albumin		3.88 ± 0.66	1.46 ± 0.04			0.05 ± 0.04	0.051 ± 0.01	
Chymotrypsinogen	4.21 ± 0.22	2.68 ± 0.14	2.36 ± 0.11	1.09 ± 0.003	0.13 ± 0.03	0.14 ± 0.03	0.14 ± 0.03	0.44 ± 0.003
RNAse	5.88 ± 0.27	4.20 ± 0.26	3.30 ± 0.15	0.23 ± 0.05	0.42 ± 0.07	0.16 ± 0.04	0.11 ± 0.02	1.26 ± 0.21
Cytochrome C	7.16 ± 0.34	4.44 ± 0.21	3.16 ± 0.14	1.78 ± 0.04	3.68 ± 0.28	0.39 ± 0.07	0.21 ± 0.04	0.37 ± 0.03
Lysozyme	5.85 ± 0.28	4.09 ± 0.21	3.54 ± 0.15	2.22 ± 0.06	1.30 ± 0.16	0.36 ± 0.07	0.30 ± 0.05	0.37 ± 0.04

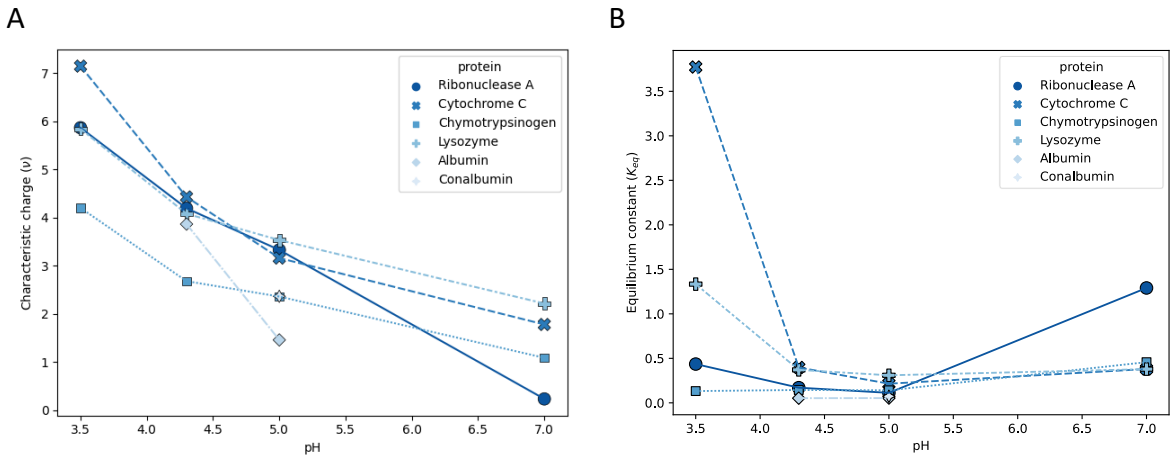


Figure 5. 3. Left: Trendlines between the (A) characteristic charge (y-axis) and (B) the equilibrium constant (y-axis) and the pH value (x-axis) for each protein.

From Table 5.3 it can observed that the characteristic charge, ν, varied between 1% and 6% of the regressed parameter value and the standard deviation values of the equilibrium constant, K_{eq}, varied between 7% and 25%. Figure 5.3A shows that the characteristic charge decreases with increasing pH for all proteins with multiple data points. This is due to the protonation of amino acids, which results in a higher net protein charge at lower pH values. A higher net charge results in more available binding sites to interact with the resin. However, no general trend can be observed between the equilibrium constant and the pH (Figure 5.3B). The equilibrium constant of cytochrome C and lysozyme decreases rapidly from pH 3.5 to pH 4.3. However, at pH 7.0 K_{eq} increases again for RNase, chymotrypsinogen, lysozyme, and cytochome C (increase of 1.19, 0.26, 0.23, and 0.23 respectively). Similar findings were reported by Yang et al., and the regressed parameters are in the same order of magnitude as reported in literature [30, 41]. In general, a higher equilibrium constant indicates a stronger binding affinity towards the resin, and therefore eluting later during the salt gradient. The same trend can be observed for the majority of proteins, see Table 5.3 and Figure 5.3. Not all proteins follow this trend, such as chymotrypsinogen, cytochrome c, and lysozyme relative to RNase (pH 7.0), and albumin relative to chymotrypsinogen (pH 4.3). These proteins elute at a later moment while having a lower equilibrium constant than the proteins eluting at an earlier moment. Though, the characteristic charge value is higher for these proteins with a lower equilibrium constant. Eventually, it is the combination of these two parameter values that determines the protein's elution moment.

5.3.1.3. Chromatographic mechanistic model validation

The chromatographic MM was validated for the gradient length of 60 CV, for pH 5.0 and 7.0. The results of pH 5.0 are shown in Figure 5.4, and of pH 7.0 in Appendix 5.C. The calibration lines convert the UV absorbance to concentration, these can be found in Appendix 5.D. As the experiments were performed in two mixtures of each three proteins, only parts of the peaks corresponding to a certain protein were used to avoid pollution of the peak by another component. In this way, the validation of each protein with the MM could be clearly evaluated.

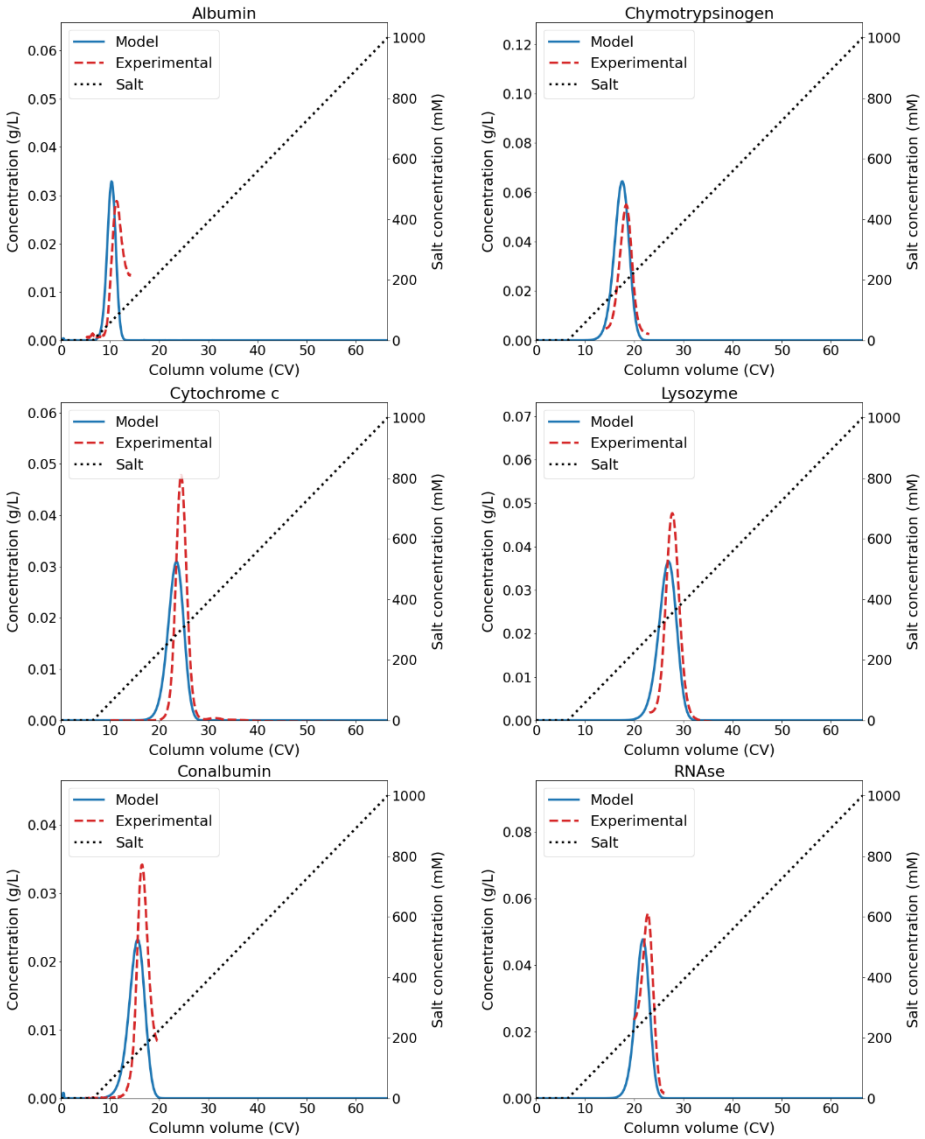


Figure 5. 4. Chromatographic mechanistic model validation for gradient length of 60 CV, equal to 58.2 mL, at a pH of 5.0. The blue line indicates the MM predicted concentration of the protein, while the red dotted line indicates the experimental concentration. The black dotted line indicates the salt concentration. The initial concentrations are albumin: 0.24 mg/mL, chymotrypsinogen: 0.80 mg/mL, conalbumin: 0.31 mg/mL, cytochrome C: 0.41 mg/mL, lysozyme: 0.55 mg/mL, and RNase: 0.56 mg/mL.

For all proteins at pH 5.0, the maximum retention peak difference is 1.04 CV and the average retention peak difference is 0.92 CV, which is 1.73% and 1.53% with respect to the gradient

length (60 CV). In all cases, except for RNase, the model predicts the start of the elution and the peak maximum earlier than the experimental results. Even though it was not be feasible to extract the entire experimental peak in all cases, it was observed that for conalbumin, cytochrome C, and lysozyme the experimental peak seems sharper than the modelled peak. To assess the concentration agreement between the modeled and experimental results, we compared the difference between the peak width at half of the peak maximum and the peak concentration. The maximum peak width difference is 1.14 CV, equal to 1.89% relative to the gradient length (60 CV). The average peak width difference is 0.81 CV, equal to 1.35% relative to the gradient length (60 CV). The average difference in the peak concentration is 0.04 mg/mL, equal to 7.36% relative to the initial concentration. Overall, the MM, using the regressed adsorption isotherm parameters, can predict the experimental data sufficiently accurate with a maximum retention peak difference of 1.73%.

5.3.2. QSPR

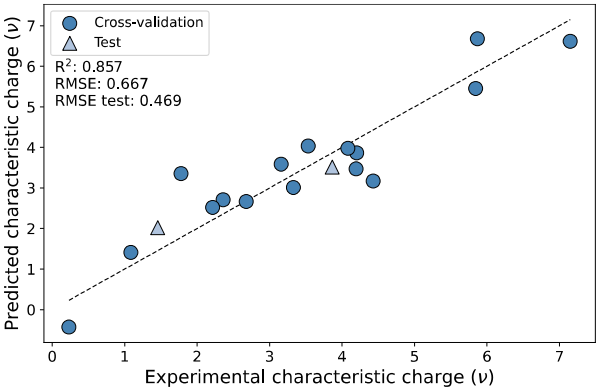
QSPR models relate specific descriptors, calculated from the protein structure, to behavior (e.g., retention). Prediction of the MM parameters, needed for simulation, starting from the protein structure allows for a full *in silico* optimization framework. From the dataset composed of the six different proteins, conalbumin at pH 5.0 was removed to be used for model verification. This protein and pH were selected because retention times for this protein were not obtained for any other pH value. This means, that conalbumin at pH 5.0 would be truly unknown for the final predictive model. The remaining 18 datapoints were split into a train and test set, where the test set was comprised of albumin measured at pH 4.3 and 5.0. As retention volumes for albumin were only obtained for pH 4.3 and 5.0, these two data points will validate the models' ability to predict the effect of differences in pH and to predict unseen proteins.

5.3.2.1. Characteristic charge

For the prediction of the characteristic charge, a MLR was trained. To avoid overfitting, a ratio of five observations to one feature should be maintained [58]. Meaning only a maximum of three features should be used in the model. To select the specific features, a redundancy filter, removing features with a Pearson correlation of >0.99 to other features, was applied. A second filter step was performed removing 40% of the features with lowest correlation to the characteristic charge. From the remaining features, sequential forward selection was performed to select the best features. A model with high accuracy (cross-validated R^2 of 0.86 and RMSE of 0.67) was obtained using only two features (Figure 5.5). As would be expected, the most important feature was related to the electrostatic potential (EP) of the protein surface. More specifically, the maximal found surface EP. The regression coefficient of this

feature was found to be 8 and permutation of the feature would result in a model not capable of predicting ν (Figure 5.5B). The second feature that was selected is the trimean of the negative hydrophobicity. This feature is less important as the regression coefficient is 1.5 and permutation results in a model with a cross-validated R^2 of 0.8. The positive regression coefficient for the second feature suggests that increasing the hydrophilicity reduces the characteristic charge. There is the possibility however, that this feature captures the titratable amino acid content on the surface, as amino acids contributing to a negative hydrophobicity are predominantly titratable. At this point we have been unable to confirm this.

A



B

Feature	Coefficient	Permutation CV R^2
Intercept	-1.26	
Maximum of surface EP	8.09	-0.52
Trimean of negative hydrophobicity	1.54	0.80

Figure 5. 5. Prediction of characteristic charge. A: Model validation of the regression model trained to predict ν where the circles represent the leave-one-out cross-validation and the triangles the test set. B: Overview of the selected features with the regression coefficient and the cross-validated R^2 after feature permutation.

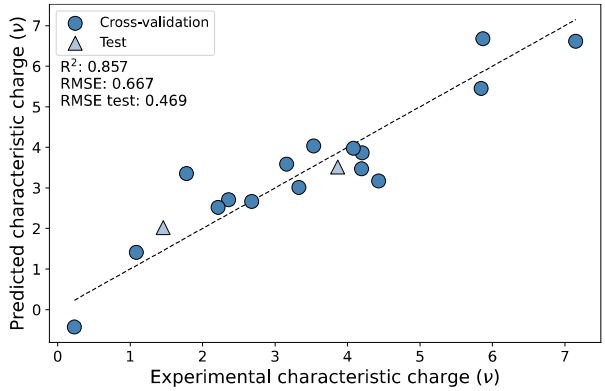
Applying the same approach to build a QSPR model for K_{eq} did not yield sufficiently accurate models. With the current dataset, the best performing models yielded only a R^2 of 0.58 (data not shown). This was considered to be insufficient for robust predictions. While ν has direct physical implications, by representing the number of charge interactions between the resin and protein, K_{eq} is lacking this [42, 59]. The equilibrium constant represents all phenomena contributing to adsorption. As observed in Figure 5.3, ν shows a clear negative trend with increasing pH, this trend is lacking for K_{eq}. It is thought that the current dataset-size is the main limitation as more features might be required to capture the complex relation. To overcome this challenge, increasing the dataset-size would result in a model trained over a greater range of property values, while also allowing an increase of the number of used features without loss of robustness [22, 30].

5.3.2.2. Retention times

Alternatively, the K_{eq} can be obtained from the regression as performed in 5.3.1.2 for experimental data. To achieve this, a MLR model for each LGE was trained (Figure 5.6). The best performing models were obtained using a feature - property correlation filter, removing

40% of the features with the lowest correlation, prior to the feature selection. The trained MLR models, for each LGE, all achieved a cross-validated R^2 of at least 0.88. For all models, the most important feature relates to the EP. More specifically, the median shell positive EP was most important for the four lower gradient lengths (20, 30, 40, and 60 CV). This feature describes the positive EP on the exterior of the protein by projecting each charge onto a plane that represents the resin. For the calculation of the shell, a total of 120 planes surround the protein, in this way representing different binding orientations. Opposed to mapping the EP onto solvent accessible surface, this method considers the distance through the solvent, penalizing protein surface within pockets. The surface fraction of alanine was the second feature selected. Alanine is a small hydrophobic amino acid, therefore this feature implicitly describes the surface hydrophobicity. The positive regression coefficient fitted for this feature indicates that a greater alanine content, and thus higher surface hydrophobicity, results in a higher retention volume. This can be explained by the salting-out effect of the Na^+ ions used during the gradient elution, resulting in hydrophobic interactions with the resin material [41].

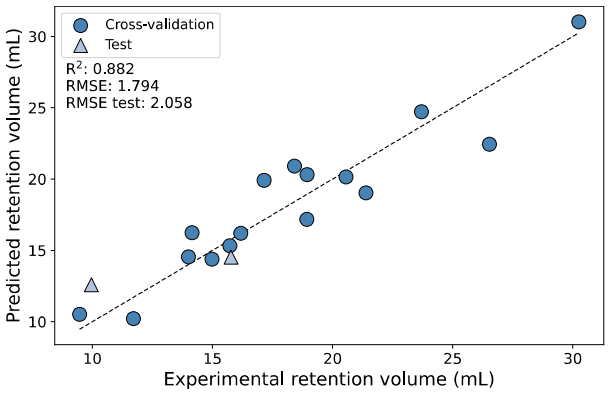
A.1



A.2

Feature	Coefficient	Permutation CV R^2
Intercept	7.47	
Median of shell positive EP	16.56	-0.17
Alanine surface fraction	2.68	0.83

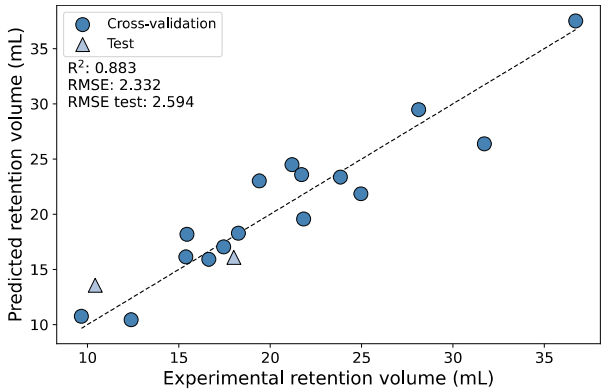
B.1



B.2

Feature	Coefficient	Permutation CV R^2
Intercept	6.50	
Median of shell positive EP	24.18	-0.18
Alanine surface fraction	4.05	0.83

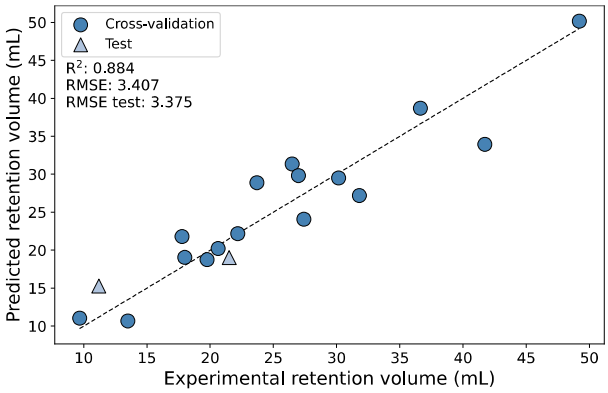
C.1



C.2

Feature	Coefficient	Permutation CV R^2
Intercept	6.39	
Median of shell positive EP	31.79	-0.20
Alanine surface fraction	5.48	0.83

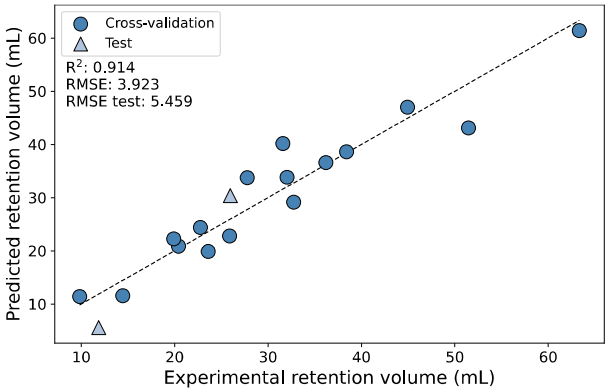
D.1



D.2

Feature	Coefficient	Permutation CV R^2
Intercept	2.97	
Median of shell positive EP	46.76	-0.21
Alanine surface fraction	8.33	0.83

E.1



E.2

Feature	Coefficient	Permutation CV R^2
Intercept	-1.74	
Mean of surface positive EP	37.73	0.85
Mean of shell positive EP	26.28	0.89
Serine surface fraction	12.76	0.83

Figure 5. 6. Prediction of protein retention at different salt gradient lengths. A to E show the validation and test of the prediction of the retention time while applying a salt gradient of 20, 30, 40, 60 and 80 column volumes, respectively. One column volume equals 0.97 mL (Table 5.1). The tables right of the plots show the feature coefficients and the effect of feature permutation on the cross-validated R^2.

For the 80 CV retention MLR model, the following features were selected: shell positive EP mean, solvent accessible surface positive EP mean, and the serine surface fraction. The feature combination yielded an accurate model with a cross-validated R^2 of 0.91 and a RMSE of 3.9 (Figure 5.6E). For the prediction of the test set, it is observed that the point at the lower end of the retention data is under predicted, compared to being over predicted in all other models. While the EP remains to be most important in the model, different features were selected during the sequential feature selection. This is due to the fact that there is no exact

linear relationship between gradient length and retention, as can be most notably observed at pH 7.0 in Appendix 5.B. While the Mean and Median of the shell EP are similar, the slight differences in the features resulted in the selection of the mean by the SFS for this model. The mean of surface positive EP and mean of shell positive EP are both important features, with regression coefficients of 37.73 and 26.28 respectively. This importance is not reflected by the permutation models, as both features describe the positive EP, collinearity allows for compensation for a loss of one of the features. However, it is essential to maintain both features to accurately predict the test set, as removing one of them results in less accurate retention estimates (data not shown). Surprisingly, the surface area fraction of serine has a positive regression coefficient, like the alanine surface fraction in the other four models. In contrast to alanine, serine is a hydrophilic residue. However, the positive regression coefficient indicates increasing retention with higher serine content on the surface, which contradicts the hypothesis for alanine selection for the previous four models. The reason behind the selection of serine in this model is currently unknown. Yet all models show good accuracy during both cross-validation and model testing, providing high confidence in model robustness.

5.3.2.3. Property prediction of conalbumin at pH 5

To demonstrate the true predictive capabilities of the trained QSPR models for the prediction of retention times and isotherm parameters, conalbumin was completely removed from the dataset prior to the train test splitting. This allowed to minimize the bias applied on the model selection. For the prediction of the retention times, the error of prediction increased with increasing gradient lengths (Table 5.4). The range of observed retention volumes rises along with the gradient lengths, likewise, the 95% confidence interval increases. Nevertheless, the effect of increasing the gradient length was captured correctly, having a maximal error of about 2 mL in retention volume, which falls within the 95% confidence interval. The characteristic charge was predicted with an error of 0.5, complying to the 95% confidence interval. Unfortunately, as no robust and accurate QSPR model for the K_{eq} could be trained with the current dataset, no direct prediction could be made. Therefore, we applied an alternative method, the predicted retention times and characteristic charge were used to regress the K_{eq}using the regression formula, similar to the experimental data method as shown in 5.3.1.2. Regression of adsorption isotherm parameters. The K_{eq} obtained was 0.028 ±0.006 which is slightly lower than the K_{eq} of 0.078 obtained by regression of the experimental data. This is due to the higher predicted ν by the QSPR model. Validation of the predicted parameters showed an accurate prediction of the conalbumin elution using a 60 CV gradient length (Figure 5.7). Both peak maximum and peak shape are simulated accurately. The difference in the peak retention time is very small, 0.12 CV, which is 0.2% difference relative to the gradient length (60 CV). The peak concentration differs by 0.009 g/L, which is

2.85% relative to the initial concentration, and the difference in the peak width at half of the peak maximum is only 1.0% relative to the gradient length (60 CV). Interestingly, the predicted parameters seem to better describe the retention profile compared to the parameters obtained from the experimental LGE, which was an average peak retention difference of 1.53% and an average peak width difference of 1.35% with respect to the gradient length (60 CV).

Table 5. 4. Predicted properties for conalbumin at pH 5.0.

Property	Experimental value (mL)	Predicted value (mL)	95% Confidence interval
Retention 20 CV	11.66	11.89	2.56
Retention 30 CV	12.89	12.92	3.69
Retention 40 CV	14.02	13.76	4.80
Retention 60 CV	16.20	15.21	7.02
Retention 80 CV	18.19	20.23	8.98
Characteristic charge (ν)	2.36	3.05	1.40

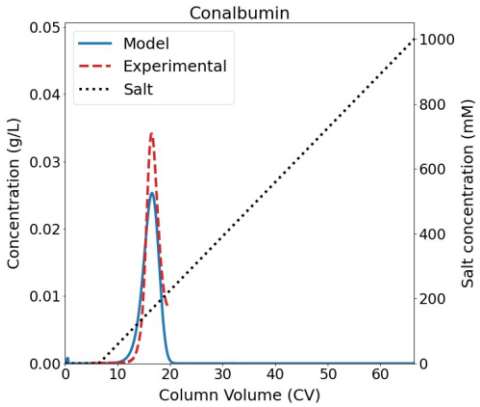


Figure 5. 7. Chromatographic mechanistic model validation of conalbumin for gradient length of 60 CV, equal to 58.2 mL, at a pH of 5.0 using the predicted isotherm parameters. Blue line indicates the MM predicted concentration of the protein, while the red dotted line indicates the experimental concentration. The black dotted line indicates the salt concentration.

5.3.3. Comparing optimization results between experimentally and QSPR-based methods

For the test protein, conalbumin at pH 5.0, both adsorption isotherm parameters, K_{eq} and ν, were determined via two methods. The first method regressed the adsorption isotherm parameters from the LGE data directly, hence LGE are needed to perform this method. While the second method involved the QSPR approach, which, after being properly trained, requires the protein-structure to determine the ν and the retention volumes. These two QSPR models were then used to regress the K_{eq} using the regression formula (Eq. 5.8).

The capture step was optimized to separate conalbumin from the other proteins, prioritizing yield over purity, utilizing the adsorption isotherm parameters determined from both

methods. This optimization aimed to assess the agreement between the optimized capture step and the parameters obtained from both methods. The resulting capture steps for both methods are depicted in Figure 5.8. The optimized variables (e.g., lower and upper cut points and the initial and final salt concentration) show comparability. The differences in both cut points are within 3.3%, and the deviation for both initial and final salt concentration is around 10 mM, approximately 3% relative to the final salt concentration (330 mM). The obtained purity only differs 0.3% and the yield 1.2% between both methods. These results demonstrate that, in this case study, it was viable to optimize the CEX capture step based solely on knowledge of the protein structure.

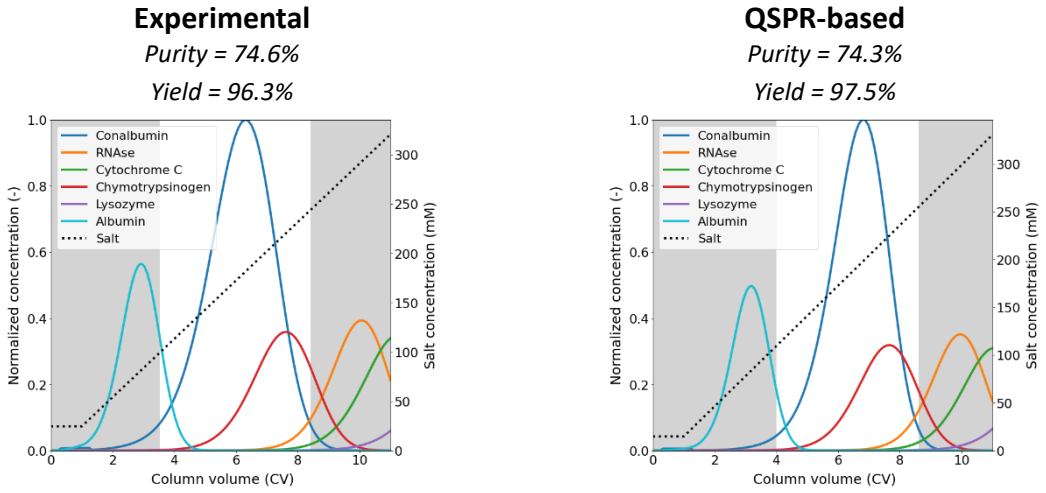


Figure 5. 8. Left: experimental-based method, the adsorption isotherm parameters were regressed directly from the LGE. K_{eq} 0.071 and ν = 2.37, lower and upper cut point are 7.7% and 91.2% respectively. The initial and final salt concentration are 24.5 mM and 320.6 mM respectively. Right: QSPR-based method, the retention volumes and ν are obtained from QSPR models, followed by using these QSPR models to regress the K_{eq} parameter. K_{eq}= 0.028 and ν = 3.05, lower and upper cut points are 4.4% and 91.7% respectively. The initial and final salt concentration are 14.8 mM and 330.4 mM respectively.

In the next part, we assessed the effect of the adsorption isotherm parameter uncertainties on the optimization outcome. We aimed to determine if variations within the standard deviation of the parameters would result in different optimal values. For both methods, numerous sample points were generated for each isotherm parameter, covering a range within their respective standard deviation. Subsequently, these sample points were used in the optimization case study. First, the consistency of the optimization case study was evaluated by running the same optimization five times, these results for both methods can be found in Appendix 5.E. This consistency evaluation aimed to ensure there were no major deviations in results within the same optimization using identical parameters. Additionally,

the minor deviations could be attributed to the optimization process itself. The optimized results for various combinations of K_{eq} and ν, ranging within their respective standard deviation, are shown in Figure 5.9 for both methods. This includes the optimized variables, such as the lower and upper cut points and the initial and final salt concentrations, as well as the purity, and the yield.

In the experimental-based method, the standard deviations for both K_{eq} (0.071±0.012) and ν (2.37±0.12) are relatively small, resulting in minimal variance in the optimized variables (Figure 5.9, A.1-F.1 and A.2-F.2, for variations in K_{eq} and ν respectively). The lower and upper cut points have a maximum difference of 7% (Figure 5.9A,B). The initial salt concentration varies between 15 and 40 mM (Figure 5.9C.1,2), and the final salt concentration is found between 320 and 327 mM (Figure 5.9D.1,2). These results suggest that despite variations in the isotherm parameters, a consistent optimum is identified, and the optimized variables exhibit only minor variations. The impact on the yield is minimal, with only a 2% variation (Figure 5.9F.1,2). On the contrary, the effect on purity is more pronounced, fluctuating between 70% and 81%. The decrease in purity is primarily attributed to an increase in the K_{eq} (Figure 5.9E.1), which is due to the greater relative standard deviation compared to ν.

For the QSPR-based method, the standard deviation of K_{eq} is small (0.028 ±0.006). The randomly spread data indicates that there is no clear correlation between K_{eq} and the optimized variables (Figure 5.9A.3-F.3). However, the standard deviation of ν is significantly larger (3.05 ±1.4), this standard deviation was defined by the 95% confidence interval calculated by Eq. 5.9. The large variation in ν resulted in two identified optima, which is clearly observed in the shift of the final salt concentration (Figure 5.6D.4). The first solution finds an optimal final salt concentration between 320 – 400 mM. The shift to the second optimal solution occurs when ν is greater than 3.6, finding the final salt concentration at around 800 mM. Remarkably, both optimal final salt concentrations are close to the set boundaries. As the characteristic charge increases, the component is expected to elute at a higher salt concentration and thus at a later moment during the gradient. This results in a greater overlap between conalbumin and the other impurities. Such a shift was not observed for the initial salt concentration, where most optimal conditions were found between 10 and 30 mM (Figure 5.9C.4). The effect of ν is also reflected in the purity and the yield (Figure 5.9E.4 and 5.9F.4 respectively). Until ν is 2.2, the purity is around 75% and the yield almost 100%, while above this value of ν, the purity increases rapidly and the yield drops to about 95%. From this point, increasing ν results in a decreasing purity and increasing yield. However, the range of the purity is broader, 50 – 85% than that of the yield, which only fluctuates between 95% and 99%. This broader range in the purity is probably due to a combination of the shift in retention

time resulting from variation of ν, and the optimization function Eq. 5.11. In the function, the yield is prioritized, representing a capture step optimization. Therefore, during challenging separation processes, the compromise on the yield is always less compared to purity. Despite the greater uncertainty in the determined v in the QSPR-method, only two optima were identified, and one of them corresponds to the optimum found in the experimental-based method.

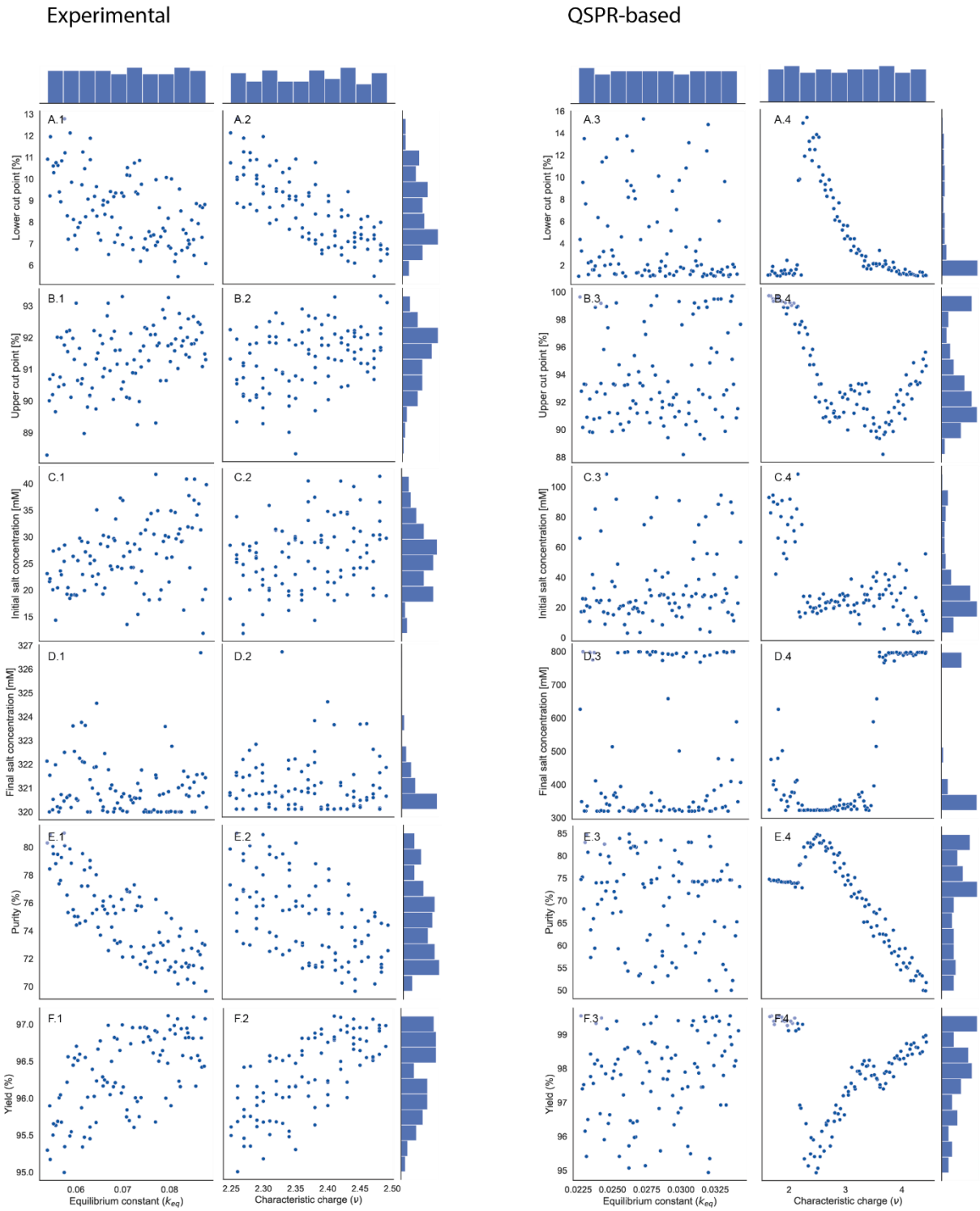


Figure 5. 9. Joint plots of scatter and hist plots between the adsorption isotherm parameters (e.g., the characteristic charge and the equilibrium constant) and the optimized variables (e.g., lower and upper cut point and the initial and final salt concentrations, and the purity and the yield). Left: experimental-based method results. Right: QSPR-based method results.

Furthermore, this optimization approach is applicable for defining the operating window of certain variables. The method employed for varying the adsorption isotherm parameters can also be used to vary other variables and assess the optimized result. In this way, the initial process design space for CPP can be defined, which is part of the QbD concept [60]. The mechanistic modeling outcomes provide knowledge on the process, therefore the number of wet-lab experiments to define the real process design space can be reduced in comparison to performing a wet-lab DoE from scratch. For the QSPR-based method, no wet-lab experiments are needed to determine the adsorption isotherm parameters and therefore the total number of experiments are even more reduced compared to the experimental-based method. For a new protein, only the protein-structure is needed to perform this optimization and make an estimation of the operating window for each optimizing variable. To illustrate, using the results from the QSPR-based method in this study, we can already narrow down the number of wet-lab DoE required to define the process design space. The final salt concentration only has to be evaluated around two main values (e.g., around 320 mM and 800 mM, see Figure 5.9D.4), while only one point of the initial salt concentration has to be assessed (e.g., 20 mM). Ultimately, the QSPR-based method offers an added advantage by allowing the incorporation of additional data over time. This not only enhances the model's accuracy, but also enables the application to other process designs, provided that the same conditions are used.

5.4. Conclusion

In this work, we demonstrated a holistic modeling approach, where we combined QSPR and chromatographic MM to optimize a CEX capture step. For an unseen protein, only the protein structure was needed to determine the adsorption isotherm parameters and predict the chromatographic retention behavior with MM. We assessed that the uncertainties in the determined adsorption isotherm parameters have a minimal and nearly equal impact for both the experimental-based and QSPR-based method.

For the experimental-based method, we successfully regressed the adsorption isotherm parameters with an R^2 minimum of 0.95. The standard deviation for the characteristic charge is within 1 – 6% of the corresponding regressed parameter value, and for the equilibrium constant, it ranges between 7 – 25% of the regressed parameter value. Moreover, the MM validation showed to be accurate with an average retention peak difference of 1.53% with respect to the gradient length.

We successfully trained MLR-QSPR models with a minimum R^2 of 0.88, even with a limited dataset composed of only five different proteins measured at four pH values. The MLR-QSPR models for predicting the characteristic charge and the retention times can be used to regress the equilibrium constant using the regression formula. A good agreement was obtained for

the MM validation for an unseen protein, conalbumin, showing only 0.2% retention peak difference with respect to the gradient length.

Both the experimental-based and the QSPR-based methods demonstrated a consistent optimized CEX capture step. The same optimum was found by both methods and an additional optimum was identified using the QSPR-based method, due to the larger standard deviation in v (3.05 ±1.4) compared to the experimentally predicted v (2.37 ±0.12). Using *in silico* optimization results as a guide can substantially reduce experimental effort, requiring experimental validation only for promising conditions. Moreover, increasing dataset sizes enhances the QSPR model accuracy, diminishing uncertainty in adsorption isotherm parameters and therefore minimizing the variance in the identified operating window.

This work highlights the value and applicability of multiscale modeling, capable to optimize a CEX capture step with only knowing the protein structure. Integrating QSPR, chromatographic MM, and optimization tools creates a versatile workflow relevant to industrial case studies. This approach enables determining initial optimal process conditions without preliminary experiments, which is especially beneficial for early phase process development when limited material and resources are available. Future applications involve extending this strategy to complex protein mixtures and broader type of chromatographic resins, offering a cost-effective and time-saving alternative that enhances overall process understanding and efficiency.

Acknowledgment

This work was partly financed from PSS-allowance for Top consortia for Knowledge and Innovation (TKI) of the ministry of Economic Affairs and partly sponsored by GlaxoSmithKline Biologicals SA 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 Vaccines and Technical University of Delft for their valuable input. Adamantia-Maria Lazopoulou was supported by the Onassis Foundation – Scholarship ID: F ZR 031-1/2021-2022.

5.5. References

[1] J.R. Birch, Y. Onakunle, Biopharmaceutical proteins: opportunities and challenges, Therapeutic proteins: Methods and protocols (2005) 1-16.

[2] E.P. Wen, R. Ellis, N.S. Pujar, Vaccine Development and Manufacturing, Wiley2014.

[3] G. Jagschies, E. Lindskog, K. Łacki, P. Galliher, Biopharmaceutical Processing: Development, Design, and Implementation of Manufacturing Processes, 2018.

[4] M. Kesik-Brodacka, Progress in biopharmaceutical development, Biotechnology and Applied Biochemistry 65(3) (2018) 306-322. https://doi.org/https://doi.org/10.1002/bab.1617.

[5] K.M. Łącki, Chapter 16 - Introduction to Preparative Protein Chromatography, in: G. Jagschies, E. Lindskog, K. Łącki, P. Galliher (Eds.), Biopharmaceutical Processing, Elsevier2018, pp. 319-366. https://doi.org/https://doi.org/10.1016/B978-0-08-100623-8.00016-5.

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

[7] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: knowledge-based chromatographic process development, Trends Biotechnol 32(4) (2014) 210-220. https://doi.org/https://doi.org/10.1016/j.tibtech.2014.02.001.

[8] I.C. Reinhardt, D.J.C. Oliveira, D.D.T. Ring, Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector, Journal of Industrial Information Integration 18 (2020) 100131. https://doi.org/https://doi.org/10.1016/j.jii.2020.100131.

[9] M. von Stosch, R.M.C. Portela, C. Varsakelis, A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice, Curr Opin Chem Eng 33 (2021) 100692. https://doi.org/https://doi.org/10.1016/j.coche.2021.100692.

[10] H. Alosert, J. Savery, J. Rheaume, M. Cheeks, R. Turner, C. Spencer, S. S. Farid, S. Goldrick, Data integrity within the biopharmaceutical sector in the era of Industry 4.0, Biotechnology Journal 17(6) (2022) 2100609. https://doi.org/https://doi.org/10.1002/biot.202100609.

[11] H. Narayanan, M.F. Luna, M. von Stosch, M.N. Cruz Bournazou, G. Polotti, M. Morbidelli, A. Butté, M. Sokolov, Bioprocessing in the Digital Age: The Role of Process Models, Biotechnology Journal 15(1) (2020) 1900172. https://doi.org/https://doi.org/10.1002/biot.201900172.

[12] A.S. Rathore, Quality by Design (QbD)-Based Process Development for Purification of a Biotherapeutic, Trends Biotechnol 34(5) (2016) 358-370. https://doi.org/https://doi.org/10.1016/j.tibtech.2016.01.003.

[13] FDA, PAT Guidance for Industry - A Framework for innovative Pharmaceutical Development, Manufacturing and Quality Assurance, 2004. www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance.

[14] ICH, ICH Harmonised Tripartite Guideline: Pharmaceutical Development Q8 (R2), ICH, 2009.

[15] J.M. Mollerup, T.B. Hansen, S. Kidal, A. Staby, Quality by design—Thermodynamic modelling of chromatographic separation of proteins, Journal of Chromatography A 1177(2) (2008) 200-206. https://doi.org/https://doi.org/10.1016/j.chroma.2007.08.059.

[16] L.K. Shekhawat, A. Tiwari, S. Yamamoto, A.S. Rathore, An accelerated approach for mechanistic model based prediction of linear gradient elution ion-exchange chromatography of proteins, Journal of Chromatography A 1680 (2022) 463423. https://doi.org/https://doi.org/10.1016/j.chroma.2022.463423.

[17] D. Saleh, G. Wang, B. Müller, F. Rischawy, S. Kluters, J. Studts, J. Hubbuch, Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications, Biotechnology Progress n/a(n/a) (2020) e2984. https://doi.org/https://doi.org/10.1002/btpr.2984.

[18] V. Kumar, A.M. Lenhoff, Mechanistic Modeling of Preparative Column Chromatography for Biotherapeutics, Annual Review of Chemical and Biomolecular Engineering 11(1) (2020) 235-255. https://doi.org/https://doi.org/10.1146/annurev-chembioeng-102419-125430.

[19] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/https://doi.org/10.1002/bit.24576.

[20] E.J. Close, J.R. Salm, D.G. Bracewell, E. Sorensen, A model based approach for identifying robust operating conditions for industrial chromatography with process variability, Chem Eng Sci 116 (2014) 284-295. https://doi.org/https://doi.org/10.1016/j.ces.2014.03.010.

[21] R. Disela, O.L. Bussy, G. Geldhof, M. Pabst, M. Ottens, Characterisation of the E. coli HMS174 and BLR host cell proteome to guide purification process development,

Biotechnology Journal 18(9) (2023) 2300068. https://doi.org/https://doi.org/10.1002/biot.202300068.

[22] D. Saleh, R. Hess, M. Ahlers-Hesse, F. Rischawy, G. Wang, J.-H. Grosch, T. Schwab, S. Kluters, J. Studts, J. Hubbuch, A multiscale modeling method for therapeutic antibodies in ion exchange chromatography, Biotechnology and Bioengineering 120(1) (2023) 125-138. https://doi.org/https://doi.org/10.1002/bit.28258.

[23] C.B. Mazza, N. Sukumar, C.M. Breneman, S.M. Cramer, Prediction of Protein Retention in Ion-Exchange Systems Using Molecular Descriptors Obtained from Crystal Structure, Analytical Chemistry 73(22) (2001) 5457-5461. https://doi.org/https://doi.org/10.1021/ac010797s.

[24] C.M. Breneman, T.R. Thompson, M. Rhem, M. Dung, Electron density modeling of large systems using the transferable atom equivalent method, Computers & Chemistry 19(3) (1995) 161-179. https://doi.org/https://doi.org/10.1016/0097-8485(94)00052-G.

[25] C.E. Whitehead, C.M. Breneman, N. Sukumar, M.D. Ryan, Transferable atom equivalent multicentered multipole expansion method, Journal of Computational Chemistry 24(4) (2003) 512-529. https://doi.org/https://doi.org/10.1002/jcc.10240.

[26] M. Song, C.M. Breneman, J. Bi, N. Sukumar, K.P. Bennett, S. Cramer, N. Tugcu, Prediction of Protein Retention Times in Anion-Exchange Chromatography Systems Using Support Vector Regression, Journal of Chemical Information and Computer Sciences 42(6) (2002) 1347-1357. https://doi.org/https://doi.org/10.1021/ci025580t.

[27] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Investigation of Mobile Phase Salt Type Effects on Protein Retention and Selectivity in Cation-Exchange Systems Using Quantitative Structure Retention Relationship Models, Langmuir 19(20) (2003) 8443-8454. https://doi.org/https://doi.org/10.1021/la0346651.

[28] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems, Proceedings of the National Academy of Sciences 102(33) (2005) 11710-11715. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.0408769102.

[29] J. Chen, S.M. Cramer, Protein adsorption isotherm behavior in hydrophobic interaction chromatography, Journal of Chromatography A 1165(1) (2007) 67-77. https://doi.org/https://doi.org/10.1016/j.chroma.2007.07.038.

[30] T. Yang, M.C. Sundling, A.S. Freed, C.M. Breneman, S.M. Cramer, Prediction of pH-Dependent Chromatographic Behavior in Ion-Exchange Systems, Analytical Chemistry 79(23) (2007) 8927-8939. https://doi.org/https://doi.org/10.1021/ac071101j.

[31] J.F. Buyel, J.A. Woo, S.M. Cramer, R. Fischer, The use of quantitative structure–activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production, Journal of Chromatography A 1322 (2013) 18-28. https://doi.org/https://doi.org/10.1016/j.chroma.2013.10.076.

[32] G. Malmquist, U.H. Nilsson, M. Norrman, U. Skarp, M. Strömgren, E. Carredano, Electrostatic calculations and quantitative protein retention models for ion exchange chromatography, Journal of Chromatography A 1115(1) (2006) 164-186. https://doi.org/https://doi.org/10.1016/j.chroma.2006.02.097.

[33] A.T. Hanke, M.E. Klijn, P.D.E.M. Verhaert, L.A.M. van der Wielen, M. Ottens, M.H.M. Eppink, E.J.A.X. van de Sandt, Prediction of protein retention times in hydrophobic interaction chromatography by robust statistical characterization of their atomic-level surface properties, Biotechnology Progress 32(2) (2016) 372-381. https://doi.org/https://doi.org/10.1002/btpr.2219.

[34] J. Kittelmann, K.M.H. Lang, M. Ottens, J. Hubbuch, Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach, Journal of Chromatography A 1510 (2017) 33-39. https://doi.org/https://doi.org/10.1016/j.chroma.2017.06.047.

[35] J. Kittelmann, K.M.H. Lang, M. Ottens, J. Hubbuch, An orientation sensitive approach in biomolecule interaction quantitative structure–activity relationship modeling and its application in ion-exchange chromatography, Journal of Chromatography A 1482 (2017) 48-56. https://doi.org/https://doi.org/10.1016/j.chroma.2016.12.065.

[36] J.R. Robinson, H.S. Karkov, J.A. Woo, B.O. Krogh, S.M. Cramer, QSAR models for prediction of chromatographic behavior of homologous Fab variants, Biotechnology and Bioengineering 114(6) (2017) 1231-1240. https://doi.org/https://doi.org/10.1002/bit.26236.

[37] R. Hess, J. Faessler, D. Yun, D. Saleh, J.-H. Grosch, T. Schwab, J. Hubbuch, Antibody sequence-based prediction of pH gradient elution in multimodal chromatography, Journal of Chromatography A 1711 (2023) 464437. https://doi.org/https://doi.org/10.1016/j.chroma.2023.464437.

[38] J. Emonts, J.F. Buyel, An overview of descriptors to capture protein properties – Tools and perspectives in the context of QSAR modeling, Computational and Structural Biotechnology Journal 21 (2023) 3234-3247. https://doi.org/https://doi.org/10.1016/j.csbj.2023.05.022.

[39] Danishuddin, A.U. Khan, Descriptors and their selection methods in QSAR analysis: paradigm for drug design, Drug Discovery Today 21(8) (2016) 1291-1302. https://doi.org/https://doi.org/10.1016/j.drudis.2016.06.013.

[40] T. Neijenhuis, O. Le Bussy, G. Geldhof, M.E. Klijn, M. Ottens, Predicting chromatographic retention of proteins using an open source QSPR workflow, Biotechnology Journal (2024), submitted for publication.

[41] Y. Hou, S.M. Cramer, Evaluation of selectivity in multimodal anion exchange systems: A priori prediction of protein retention and examination of mobile phase modifier effects, Journal of Chromatography A 1218(43) (2011) 7813-7820. https://doi.org/https://doi.org/10.1016/j.chroma.2011.08.080.

[42] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, J. Hubbuch, Determination of parameters for the steric mass action model—A comparison between two approaches, Journal of Chromatography A 1233 (2012) 54-65. https://doi.org/https://doi.org/10.1016/j.chroma.2012.02.004.

[43] F. Hagemann, P. Adametz, M. Wessling, V. Thom, Modeling hindered diffusion of antibodies in agarose beads considering pore size reduction due to adsorption, Journal of Chromatography A 1626 (2020) 461319. https://doi.org/https://doi.org/10.1016/j.chroma.2020.461319.

[44] D. Keulen, E. van der Hagen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Using artificial neural networks to accelerate flowsheet optimization for downstream process development, Biotechnology and Bioengineering (2023) 1-14. https://doi.org/https://doi.org/10.1002/bit.28454.

[45] D.M. Ruthven, Principles of adsorption and adsorption processes, John Wiley & Sons, New York, 1984.

[46] L. Petzold, Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations, SIAM Journal on Scientific and Statistical Computing 4(1) (1983) 136-148. https://doi.org/https://doi.org/10.1137/0904010.

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

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

[49] S.M. Pirrung, D.P. da Cruz, A.T. Hanke, C. Berends, R.F.W.C. van Beckhoven, M.H.M. Eppink, M. Ottens, Chromatographic Parameter Determination for Complex Biological Feedstocks, Biotechnology Progress 34(4) (2018) 1006-1018. https://doi.org/10.1002/btpr.2642.

[50] T. Hahn, N. Geng, K. Petrushevska-Seebach, M.E. Dolan, M. Scheindel, P. Graf, K. Takenaka, K. Izumida, L. Li, Z. Ma, N. Schuelke, Mechanistic modeling, simulation, and optimization of mixed-mode chromatography for an antibody polishing step, Biotechnology Progress 39(2) (2023) e3316. https://doi.org/https://doi.org/10.1002/btpr.3316.

[51] E.S. Parente, D.B. Wetlaufer, Relationship between isocratic and gradient retention times in the high-performance ion-exchange chromatography of proteins: Theory and experiment, Journal of Chromatography A 355 (1986) 29-40. https://doi.org/https://doi.org/10.1016/S0021-9673(01)97301-7.

[52] H. Schmidt-Traub, M. Schulte, A. Seidel-Morgenstern, H. Schmidt-Traub, Preparative chromatography, Wiley Online Library2012.

[53] A.A. Shukla, S.S. Bae, J.A. Moore, K.A. Barnthouse, S.M. Cramer, Synthesis and Characterization of High-Affinity, Low Molecular Weight Displacers for Cation-Exchange Chromatography, Ind Eng Chem Res 37(10) (1998) 4090-4098. https://doi.org/https://doi.org/10.1021/ie9801756.

[54] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Research 28(1) (2000) 235-242. https://doi.org/https://doi.org/10.1093/nar/28.1.235.

[55] J. Rodrigues, J. Teixeira, M. Trellet, A. Bonvin, pdb-tools: a swiss army knife for molecular structures [version 1; peer review: 2 approved], F1000Research 7(1961) (2018). https://doi.org/https://doi.org/10.12688/f1000research.17456.1.

[56] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, Journal of Chemical Theory and Computation 7(2) (2011) 525-537. https://doi.org/https://doi.org/10.1021/ct100578z.

[57] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, Journal of

Computational Chemistry 25(13) (2004) 1605-1612. https://doi.org/https://doi.org/10.1002/jcc.20084.

[58] J.G. Topliss, R.J. Costello, Chance correlations in structure-activity studies using multiple regression analysis, Journal of Medicinal Chemistry 15(10) (1972) 1066-1068. https://doi.org/https://doi.org/10.1021/jm00280a017.

[59] C.A. Brooks, S.M. Cramer, Steric mass-action ion exchange: Displacement profiles and induced salt gradients, AIChE Journal 38(12) (1992) 1969-1978. https://doi.org/https://doi.org/10.1002/aic.690381212.

[60] A.S. Rathore, Roadmap for implementation of quality by design (QbD) for biotechnology products, Trends Biotechnol 27(9) (2009) 546-553. https://doi.org/https://doi.org/10.1016/j.tibtech.2009.06.006.

Chapter 6

Conclusions and Outlook

6.1. Conclusions

This work presents original research in the field of modeling and optimization of biopharmaceutical downstream processes. The developed chromatographic mechanistic modeling software enables to simulate different modes of chromatography and is flexible in adapting the mass transfer correlations, hydrodynamics and adsorptions isotherms. The integration of several mechanistic models allows to optimize the entire downstream process *in silico*, including chromatography and ultrafiltration / diafiltration steps. Moreover, we investigated the use of different modeling techniques, ANNs and mechanistic models, and optimization strategies to determine the most effective approach for complex flowsheet optimizations. Lastly, to reduce experimental efforts for determining adsorption isotherm parameters, QSPR modeling is applied in combination with mechanistic modeling to optimize a capture step. The general conclusions of the research topics in this work are summarized in the following.

In **Chapter 2**, a comprehensive overview is given of the present and future downstream process development strategies and tools utilized in the (bio)pharmaceutical industry and academia. The following conclusions are drawn:

- The vaccine purification process development is highly experimentally driven. This highlighted the need for modernizing strategies in (protein subunit) vaccine process development, such as establishing a standardized approach or platform process, and enhancing the understanding of host cell impurities.
- Modeling techniques can play a crucial role in reducing experimental effort and enhancing process understanding. The combination of diverse modeling techniques will advance the implementation of model-based process development approaches by mitigating the limitations associated with each individual modeling technique.
- High Throughput Process Development is crucial for reducing the consumption of resource materials while allowing for extensive exploration of a large design space, particularly in the early stages of process development when product materials are limited.

The assessment in **Chapter 3** focused on employing ANNs instead of chromatographic mechanistic modeling during flowsheet optimization. This approach aimed to achieve greater time efficiency while still identifying the most optimal sequences. In summary, the following insights are gained:

- For a case study considering a maximum of three different chromatography operations in a sequence, it was demonstrated that using ANNs decreased the overall optimization time by 50%.
- Based on the outcome of the global optimization, the most promising flowsheets can be pre-selected. This substantially reduces the number of flowsheets to be optimized during local optimization, resulting in a significant reduction in overall optimization time.
- ANNs prove to be effective for global optimization, enabling decision-making on type and number of unit operations.

In **Chapter 4**, we compared three optimization strategies during flowsheet optimization in terms of outcome, complexity, and time-efficiency. Each strategy (e.g., simultaneous, top-to-bottom, and superstructure decomposition) is solved using both mechanistic models and ANNs to compare the influence of each modeling technique. This analysis leads to the following conclusions:

- The overall weighted performance values, predicted within a similar range for both modeling techniques (e.g., mechanistic models and ANNs), confirm the accuracy of ANNs for complex flowsheet optimization, as discussed in Chapter 3.
- All optimization strategies identified similar optimal flowsheets, each consisting of three steps and an orthogonal structure. In these optimized structures, salt conditions are adjusted to prioritize a dilution step over a diafiltration mode.
- The superstructure decomposition method with MM is the most time-efficient. It enables to complete the optimization in less than 40 computational hours when utilizing multiple cores, meaning that a computer, containing multiple processing cores, can perform multiple simulations simultaneously.

A multiscale modeling approach is presented in **Chapter 5**, where we combined QSPR and chromatographic mechanistic modeling techniques. This study yields the following key takeaways:

- Through QSPR modeling and the regression formula, we obtained the adsorption isotherm parameters by only knowing the protein structure. With these *in silico* predicted isotherm parameters, the chromatographic retention behavior can be predicted using mechanistic modeling. The results demonstrate a strong agreement with the experimental data, revealing only 0.2% difference in retention peak values relative to the gradient length.

- An assessment was conducted to evaluate how the variability in adsorption isotherm parameters, determined through both experimental-based and QSPR-based methods, affects the optimization outcome. Both the experimental-based and the QSPR-based methods revealed a consistently optimized CEX capture step. The same optimum is found by both methods, the QSPR-based method identified an additional optimum due to a higher standard deviation in one of the isotherm parameters.
- This multiscale modeling approach highlights the substantial reduction in experimental efforts achieved through *in silico* optimization. This enables the determination of initial optimal process conditions without the need for preliminary experiments.

Due to an increasing population, intensified international traveling, and the resistance against antibiotics, emerging infectious diseases can even spread faster and become more harmful. In response, it is crucial to modernize the vaccine process development. The goal is to design a process within a short time frame that is efficient, robust and scalable for large-scale vaccine production. This thesis emphasizes the added value of modeling techniques in process development. As model-based approaches reduce experimental effort, enhance process understanding, and enable to screen the overall design space. This work is especially valuable for early phase process development when limited material and resource are available.

6.2. Outlook

For future prospect, several areas related to the modeling and optimization of biopharmaceutical downstream processes remain interesting for further exploration and advancement.

Coupling the upstream and downstream processes *in silico*, using detailed mechanistic models, would enhance our understanding of the entire integrated process and enable optimization of the overall process. Moreover, this is interesting for advanced control strategies throughout the entire process, particularly in the context of continuous biomanufacturing as done by Gomis-Fons et al. for a mAb production process [1]. Recently, Wahlgreen et al. presented a numerical case study for the production of a mAb, where the fed batch reactor was connected to a chromatography capture step [2]. Although, this numerical connection of unit operations gives additional insight information, experimental validation is still lacking. Simultaneously optimizing the complete process *in silico*, followed by experimentally validating the optimized outcome, would be very valuable. This approach could substantiate the applicability and, ideally, the precision of modeling applied in process

development. Simultaneous optimization of upstream and downstream processes could balance the upstream-yield with the impurities produced that are consequently hard to remove downstream.

Furthermore, it would be interesting to integrate the detailed mechanistic models with a process modeling software, such as gPROMS [3], AspenTech [4] or SuperPro Designer [5]. Particularly gPROMS and AspenTech software are more based on chemical processes and therefore lacking specifications or advanced options for biopharmaceutical processes. The advantage of process modeling software is the ability to model and directly visualize the complete process. It allows to easily adapt the connections between unit operations and provides flexibility to choose from various options. Although SuperPro Designer has more possibilities for biopharmaceutical processes compared to the other software, the integrated models are limited and primarily based on simplified mass or energy balances [6]. Integrating the detailed mechanistic models with available process modeling software would get the best out of both applications. As it reduces the complexity to model the entire process, while the details of the models are retained. A bottleneck is that these process modeling software are only commercially available. A follow-up step would be to make an (flowsheet) optimization code around the process modeling software. Another option would be to integrate the process modeling software with superstructure generation software such as super-O [7], P-graph [8], or Pyosyn [9]. Adding an additional optimization layer poses a greater challenge and potentially slows down the optimization process when using different software. However, transforming this software into a more user-friendly software or combined with other available user-friendly software will increase its usability.

For this work, we used a pre-defined generated superstructure, for which all possibilities were evaluated as our case studies did not involve that many different unit operations. For even more complex flowsheets involving additional constraints, such as considering seven different types of unit operations with specific restrictions on their positions in the sequence, creating a pre-defined generated superstructure can omit certain process options. For these reasons, exploring superstructure-free approaches, such as reinforcement learning, would be interesting [10, 11]. This method begins with one unit operation and then employs a random search for the next one. It makes decisions based on the outcome of each unit operation, determining whether to proceed or discontinue with a specific sequence. Ultimately, the goal is to find the optimal process through this iterative approach [12].

The decomposition of the superstructure allows to apply different objectives to parts of the downstream process and would be interesting to study (Chapter 3). For example, applying a higher weight on the yield for the capture step, while making the purity more important

during the polishing steps. Moreover, the way of formulating the multi-objective optimization could be further explored. In this work, we assumed the weights in the multi-objective in advance. However, to find a solution that optimally balances the conflicting objectives, a multiple-criteria decision making method can be employed. This approach involves determining the optimal weights in the multi-objective function through the use of an algorithm [13, 14]. In addition, the type of optimization solvers to perform the optimization can also be studied into more depth. This could potentially make the optimization more time-efficient.

Combining different modeling techniques can bring great benefits, such as the multiscale modeling approach as presented in Chapter 5. As a follow-up, this multiscale modeling approach can be applied to a complex mixture considering numerous proteins using mass spectrometry as analytical technique, to interrogate adsorption and retention behavior. This database can be used to train, validate, and test QSPR models. Alternatives for QSPR models can also be investigated, such as Graph Neural Networks, which belongs to the class of ANNs [15]. Moreover, in this work a Deep Neural Network, which is an ANN with multiple hidden layers, was used as surrogate model for the mechanistic model. However, it would be advantageous if the same or improved accuracy can be achieved requiring less data, therefore, it would be interesting to explore various ANN classes. Combinatorial modeling approaches that may be interesting to further study depend on the mechanistic model's applications. For example, when creating a digital twin, exploring hybrid modeling is of interest to improve the accuracy [16]. While physics informed neural networks could be relevant if parameters are unknown or to develop a reduced order model [17].

Furthermore, as modeling applications and big data are emerging rapidly, the efficient use and sustainability of software, as well as the efficient processing of big data, have become crucial. Therefore, standardizing code-writing would be beneficial, especially when collaborating within a project or group. Platforms like GitLab or GitHub are convenient tools for sharing and collaborating on codes. This also applies for processing big data, a structured code that is easily accessible via a platform is beneficial for everyone involved. This demands for a structural organization of software/codes and involvement of software engineers.

6.3. References

[1] J. Gomis-Fons, H. Schwarz, L. Zhang, N. Andersson, B. Nilsson, A. Castan, A. Solbrand, J. Stevenson, V. Chotteau, Model-based design and control of a small-scale integrated continuous end-to-end mAb platform, Biotechnology Progress 36(4) (2020) e2995. https://doi.org/https://doi.org/10.1002/btpr.2995.

[2] M.R. Wahlgreen, K. Meyer, T.K.S. Ritschel, A.P. Engsig-Karup, K.V. Gernaey, J.B. Jørgensen, Modeling and Simulation of Upstream and Downstream Processes for Monoclonal Antibody Production, IFAC-PapersOnLine 55(7) (2022) 685-690. https://doi.org/https://doi.org/10.1016/j.ifacol.2022.07.523.

[3] gPROMS Digital Process Design and Operations. https://www.siemens.com/global/en/products/automation/industry-software/gproms-digital-process-design-and-operations.html.

[4] A.T. Inc., AspenTech. https://www.aspentech.com/en/products/engineering/aspen-plus.

[5] I. SuperPro Intelligen, SuperPro Designer. https://www.intelligen.com/products/superpro-overview/.

[6] D. Petrides, D. Carmichael, C. Siletti, A. Koulouris, Biopharmaceutical process optimization with simulation and scheduling tools, Bioengineering 1(4) (2014) 154-187.

[7] M.-O. Bertran, R. Frauzem, L. Zhang, R. Gani, A Generic Methodology for Superstructure Optimization of Different Processing Networks, in: Z. Kravanja, M. Bogataj (Eds.), Computer Aided Chemical Engineering, Elsevier2016, pp. 685-690. https://doi.org/https://doi.org/10.1016/B978-0-444-63428-3.50119-3.

[8] F. Friedler, K.B. Aviso, B. Bertok, D.C.Y. Foo, R.R. Tan, Prospects and challenges for chemical process synthesis with P-graph, Curr Opin Chem Eng 26 (2019) 58-64. https://doi.org/https://doi.org/10.1016/j.coche.2019.08.007.

[9] Q. Chen, Y. Liu, G. Seastream, J.D. Siirola, I.E. Grossmann, Pyosyn: A new framework for conceptual design modeling and optimization, Comput Chem Eng 153 (2021) 107414. https://doi.org/https://doi.org/10.1016/j.compchemeng.2021.107414.

[10] Q. Gao, A.M. Schweidtmann, Deep reinforcement learning for process design: Review and perspective, arXiv preprint arXiv:2308.07822 (2023).

[11] S. Nikita, A. Tiwari, D. Sonawat, H. Kodamana, A.S. Rathore, Reinforcement Learning based Optimization of Process Chromatography for Continuous Processing of

Biopharmaceuticals, Chem Eng Sci (2020) 116171. https://doi.org/https://doi.org/10.1016/j.ces.2020.116171.

[12] L. Wang, M. Lampe, P. Voll, Y. Yang, A. Bardow, Multi-objective superstructure-free synthesis and optimization of thermal power plants, Energy 116 (2016) 1104-1116. https://doi.org/https://doi.org/10.1016/j.energy.2016.10.007.

[13] R.T. Marler, J.S. Arora, Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization 26(6) (2004) 369-395. https://doi.org/10.1007/s00158-003-0368-6.

[14] I.Y. Kim, O.L. de Weck, Adaptive weighted sum method for multiobjective optimization: a new method for Pareto front generation, Structural and Multidisciplinary Optimization 31(2) (2006) 105-116. https://doi.org/10.1007/s00158-005-0557-6.

[15] A.M. Schweidtmann, J.G. Rittig, A. König, M. Grohe, A. Mitsos, M. Dahmen, Graph Neural Networks for Prediction of Fuel Ignition Quality, Energy & Fuels 34(9) (2020) 11395-11407. https://doi.org/10.1021/acs.energyfuels.0c01533.

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

[17] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686-707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

Supplementary material

S

Appendix - Chapter 3

Appendix 3.A

Table 3.A.1. Overview of required Python libraries

Library	Version
python	3.8.5
scipy	1.7.3
numpy	1.19.2
spyder	4.1.5
pandas	1.2.3
matplotlib	3.3.4
openpyxl	3.0.7
notebook	6.2.0
Ipywidgets	7.6.3
Tensorflow	2.10.1
Keras-appplications	2.10.0
Dataclasses	0.8

Appendix 3.B

We used the linear multicomponent mixed-mode isotherm, as formulated by by Nfor et al. [1], to calculate the equilibrium concentration in the liquid phase:

$$\frac{q_i}{C_i} = K_{eq,i}\Lambda^{(v_i+n_i)}(z_s c_s)^{-v_i} {c_v}^{-n_i}\gamma_i \qquad \textit{Eq. 3.A.1}$$

where subscripts i denotes the protein component. n_i is the hydrophobic interaction stoichiometric coefficient and v_i is the stoichiometric coefficient of salt counter ions, calculated by dividing the effective binding charge of the protein (z_p) with the charge on the salt counter ion (z_s), $v_i = z_p/z_s$. Λ is the ligand density of the mixed mode resins, c_s is the salt concentration in the liquid phase, c_v is the molarity of the solution in the pore volume, and γ is the activity coefficient of the protein solution. Often, either one of the interaction modes is dominant and therefore the equation can be simplified by setting $n = 0$, in case of ion-exchange chromatography and $v = 0$ for HIC chromatography. Details of the isotherm and resin parameters for each chromatography mode and all proteins are given in Table 6. The bed porosity was assumed to be 0.27 and the total porosity 0.95.

Table 3.B.1. Details of the isotherm and resin parameters used for the chromatography model [2], protein 1 = mAb, protein 2 = Moesin, protein 3 = Chitotrisidase, protein 4 = Legumain, and protein 5 = Thioredoxin reductase.

	CEX	AEX	HIC
Resin	Mono S	Mono Q	Source PHE
Particle diameter (µm)	30	30	15

Pore diameter (nm)		40	40	83.4
Ligand density (Λ)		135	320	40
Column volume		1 mL	1 mL	1 mL
Flowrate		150 cm/h	150 cm/h	150 cm/h
Keq	Protein 1	8.5	0.5	9.3
	Protein 2	500.8	0.5	1.6
	Protein 3	604.2	0.9	10.4
	Protein 4	0.0	3.9	9.3
	Protein 5	8.5	3.9	1.6
$\boldsymbol{v}$ $(\boldsymbol{or}$ $\boldsymbol{n})$	Protein 1	2.6	4.0	9.3
	Protein 2	2.5	4.0	1.6
	Protein 3	2.6	1.7	10.4
	Protein 4	0.0	2.9	9.3
	Protein 5	2.6	2.9	1.6

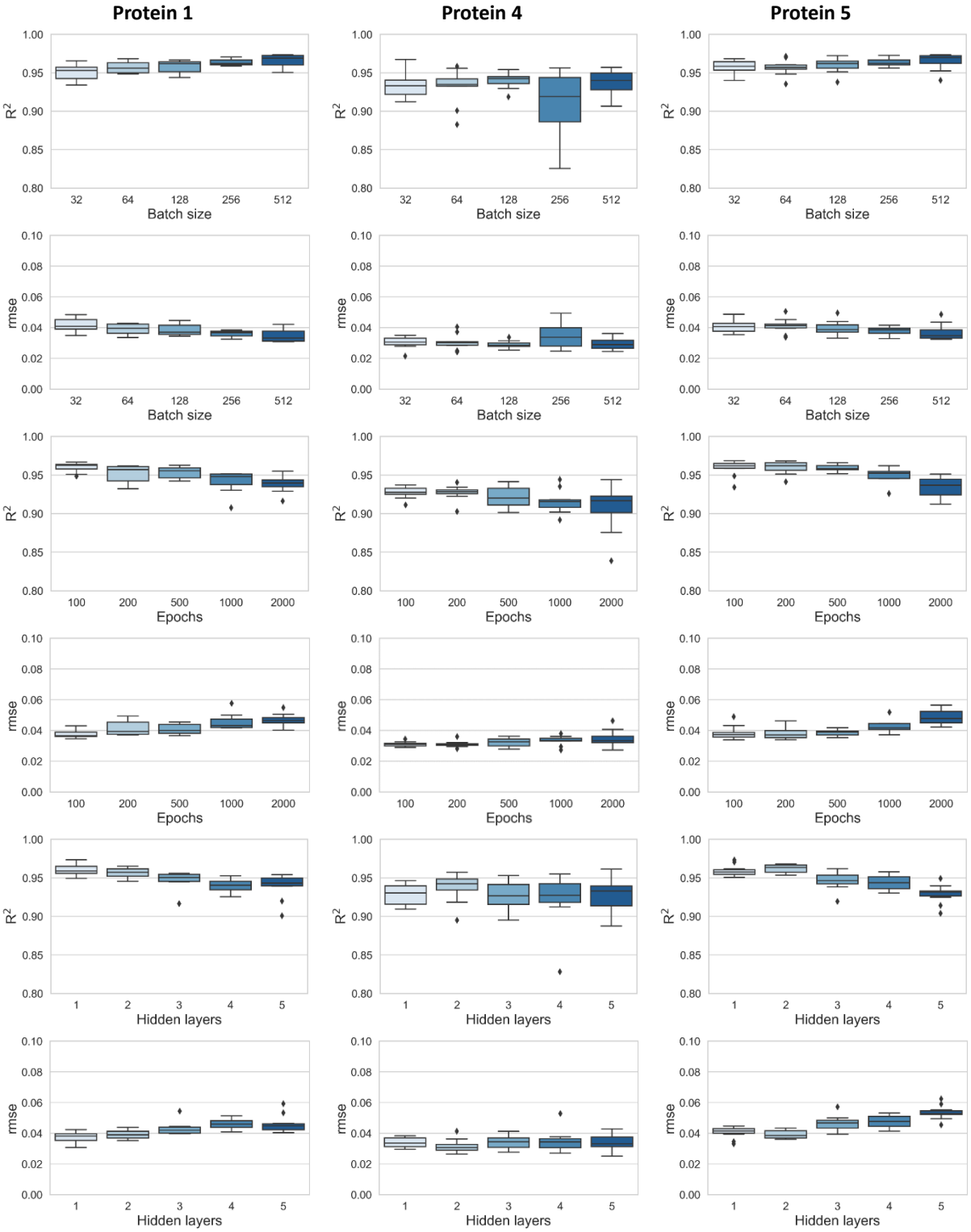
[1] B.K. Nfor, M. Noverraz, S. Chilamkurthi, P.D.E.M. Verhaert, L.A.M. van der Wielen, M. Ottens, High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents, Journal of Chromatography A 1217(44) (2010) 6829-6850. https://doi.org/https://10.1016/j.chroma.2010.07.069.

[2] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/https://doi.org/10.1002/bit.24576

Appendix 3.C

Hyperparameters evaluations (e.g., batch size, and the number of epochs, hidden layers, and neurons) for each chromatography mode (e.g., CEX, AEX, and HIC).

CEX



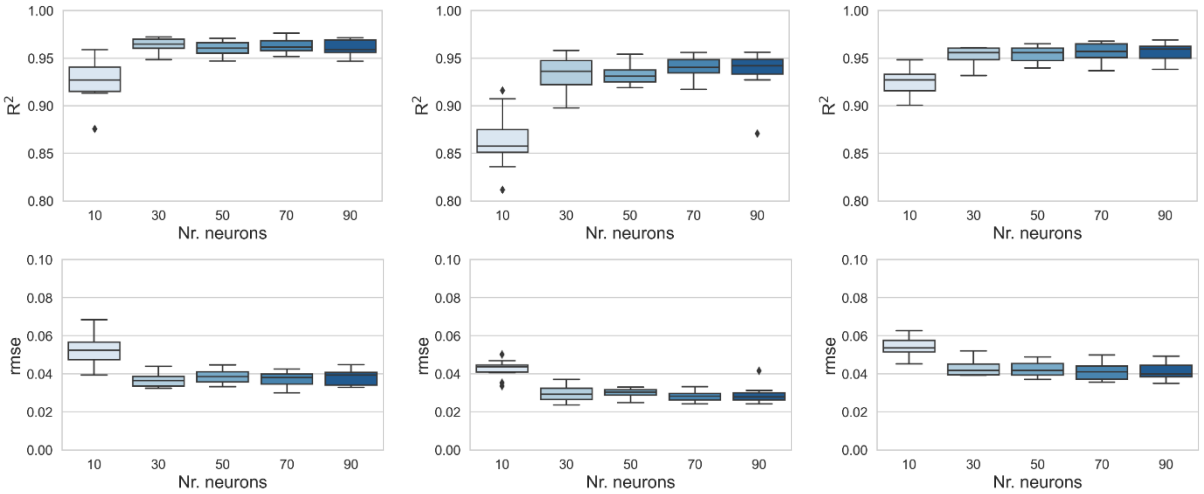


Figure 3.C.1. Boxplots of CEX for proteins 1, 4, and 5, showing the accuracy (R^2 and RMSE) for several hyperparameters (e.g., batch size, and the number of epochs, hidden layers, and neurons). Proteins 2 and 3 were not considered for the hyperparameter evaluation as these proteins were never present in the product and therefore showed a very low R^2.

S

AEX

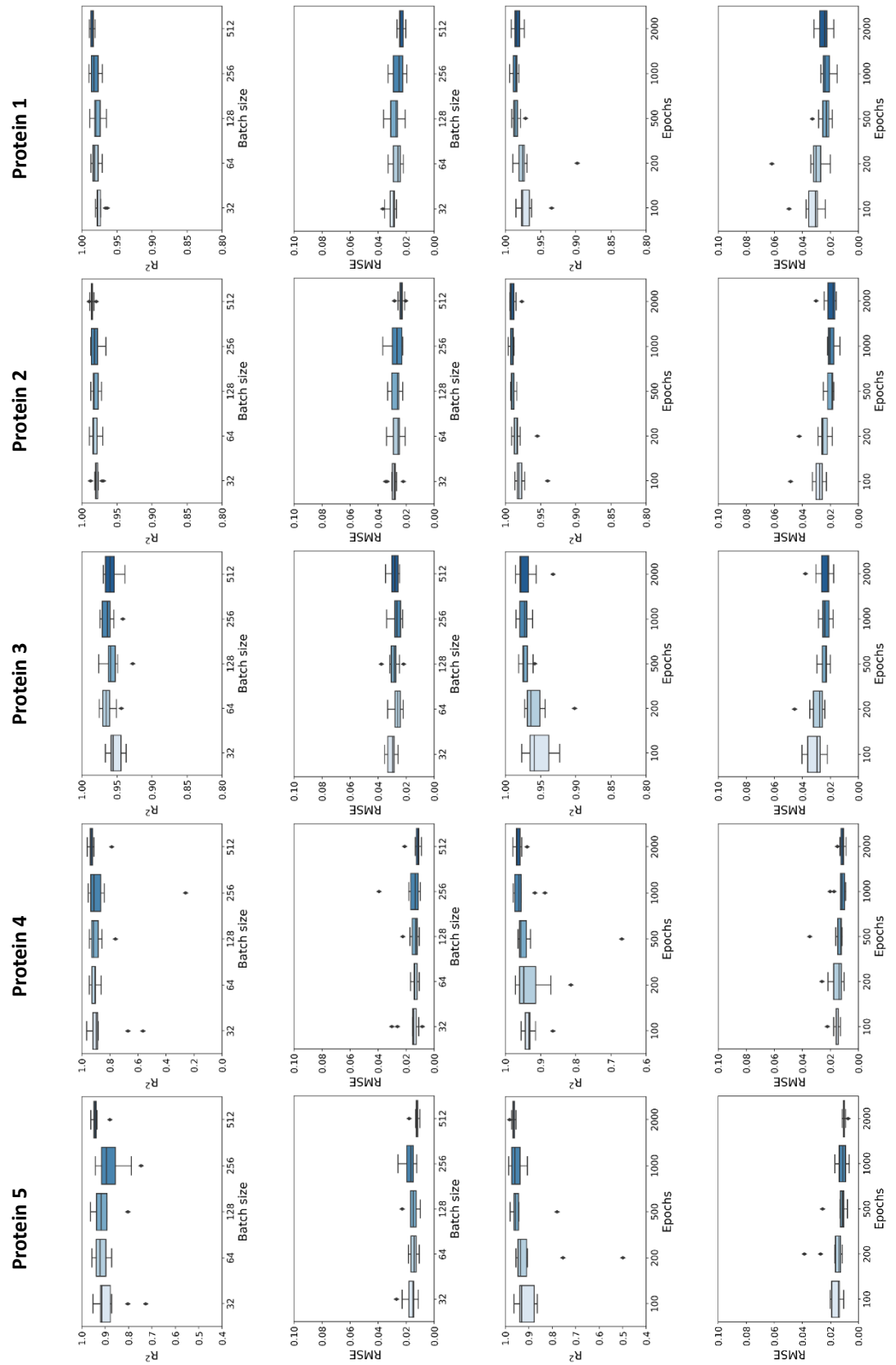


Figure 3.C.2. Boxplots of AEX for proteins 1, 2, 3, 4, and 5, showing the accuracy (R^2 and RMSE) for several hyperparameters (e.g., batch size, and the number of epochs, hidden layers, and neurons).

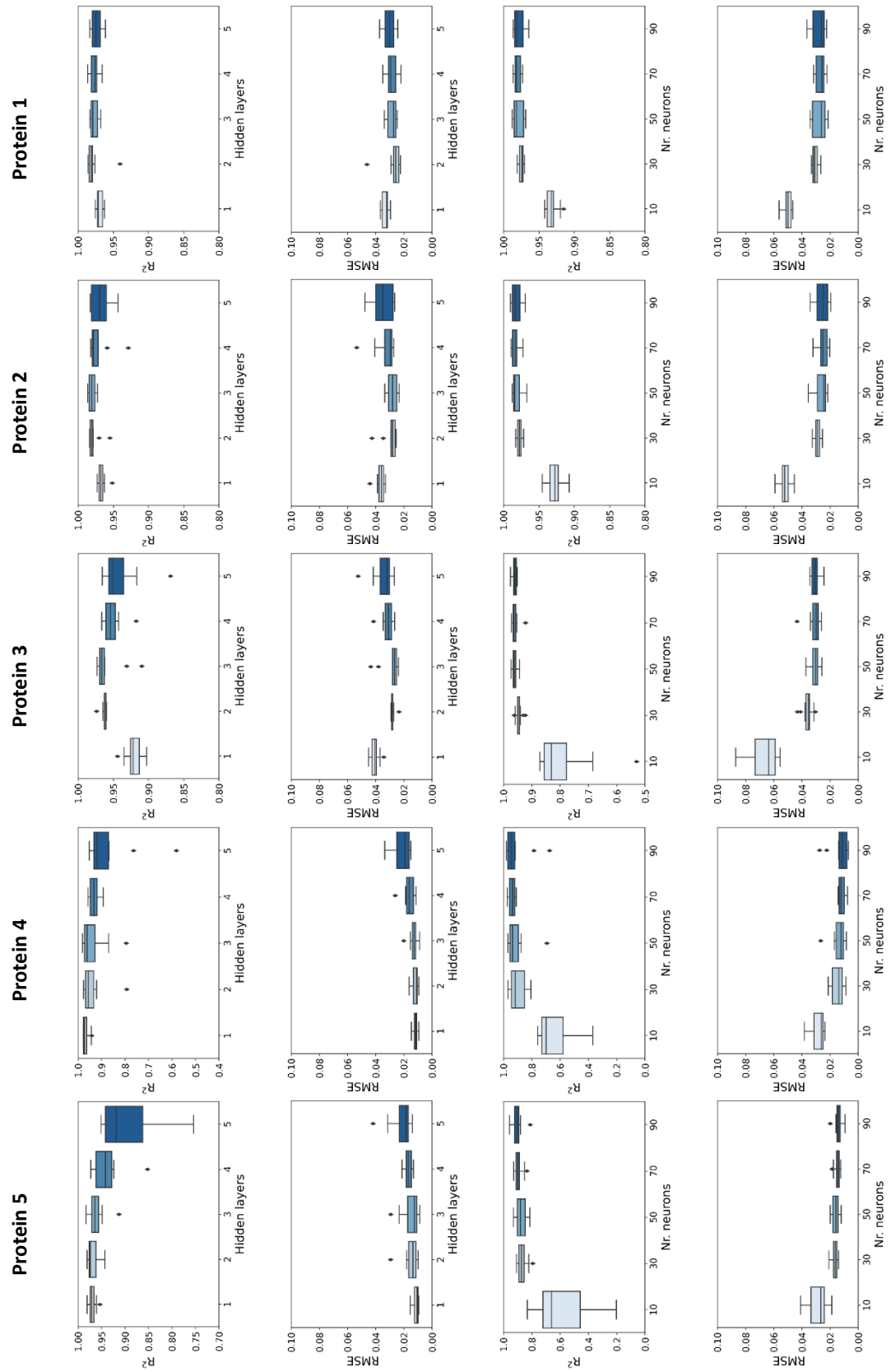


Figure 3.C.2. Continuation, boxplots of AEX for proteins 1, 2, 3, 4, and 5, showing the accuracy (R^2 and RMSE) for several hyperparameters (e.g., batch size, and the number of epochs, hidden layers, and neurons).

HIC

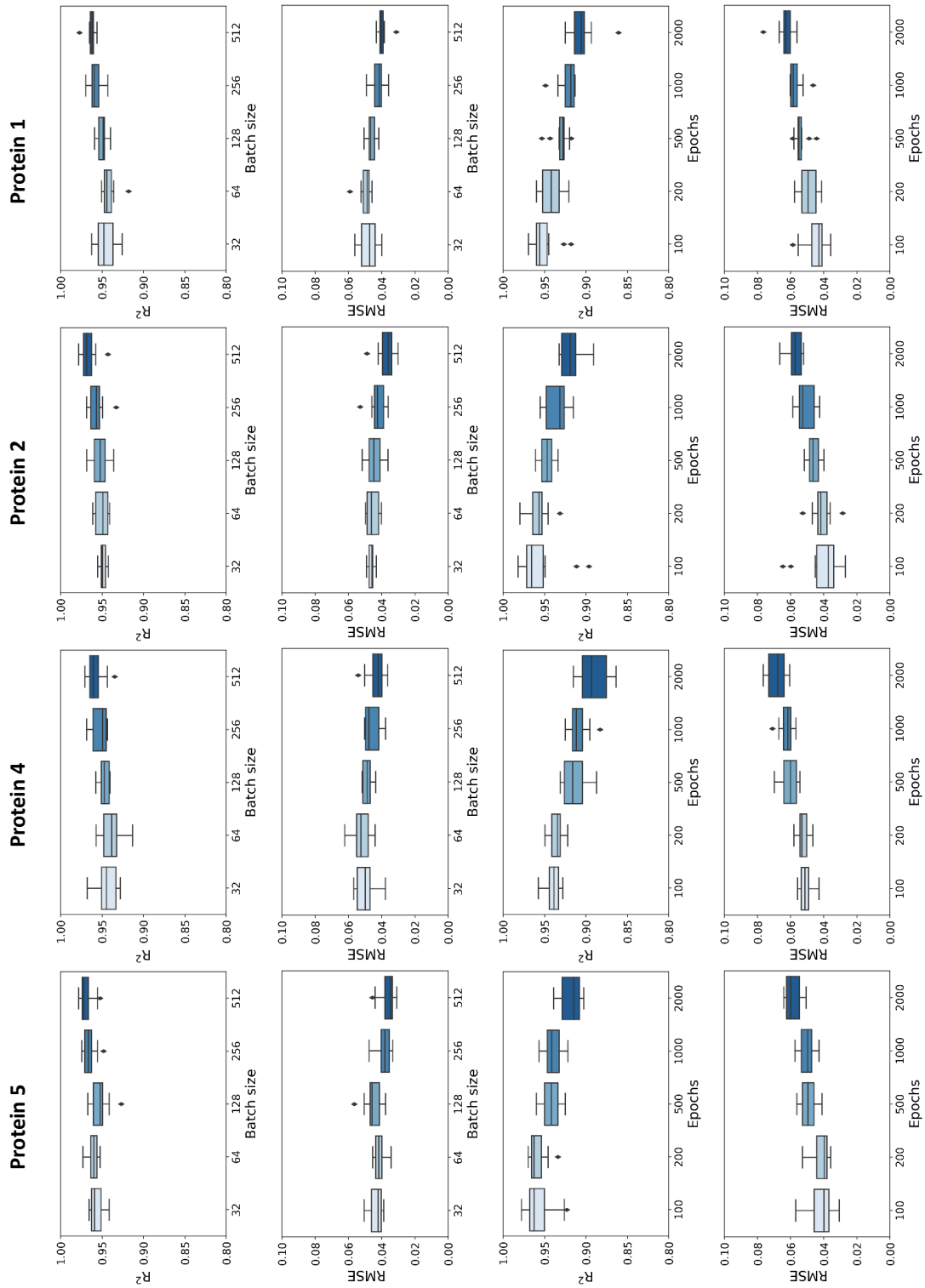


Figure 3.C.3. Boxplots of HIC for proteins 1, 2, 4, and 5, showing the accuracy (R^2 and RMSE) for several hyperparameters (e.g., batch size, and the number of epochs, hidden layers, and neurons). Protein 3 was not considered for the hyperparameter evaluation as this proteins were never present in the product and therefore showed a very low R^2.

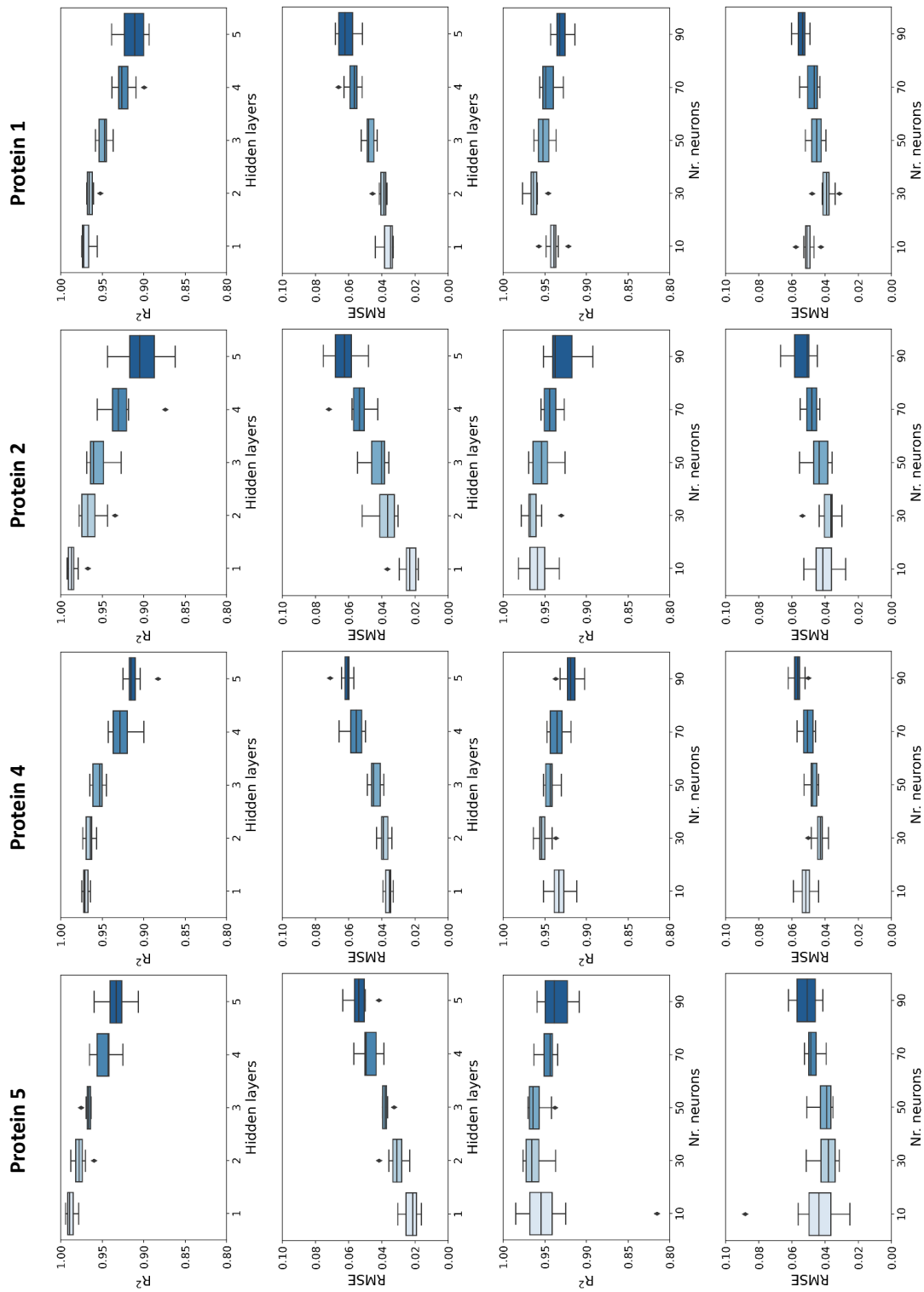


Figure 3.C.3. Continuation, boxplots of HIC for proteins 1, 2, 4, and 5, showing the accuracy (R2 and RMSE) for several hyperparameters (e.g., batch size, and the number of epochs, hidden layers, and neurons). Protein 3 was not considered for the hyperparameter evaluation as this proteins was never present in the product and therefore showed a very low R2.

S

Appendix 3.D

CEX

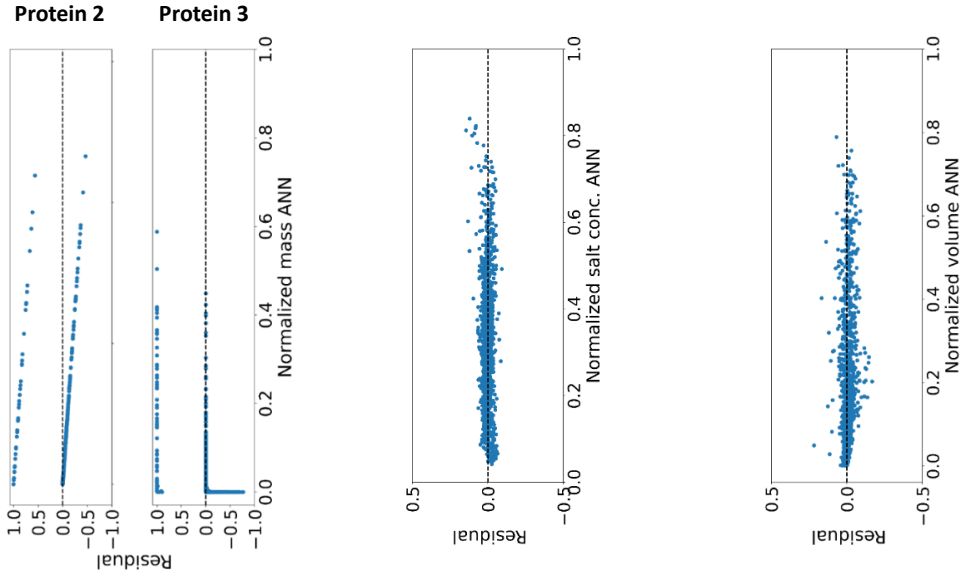


Figure 3.D.1. Residual figures of the CEX mode for proteins 2 and 3, and the salt concentration and volume.

AEX

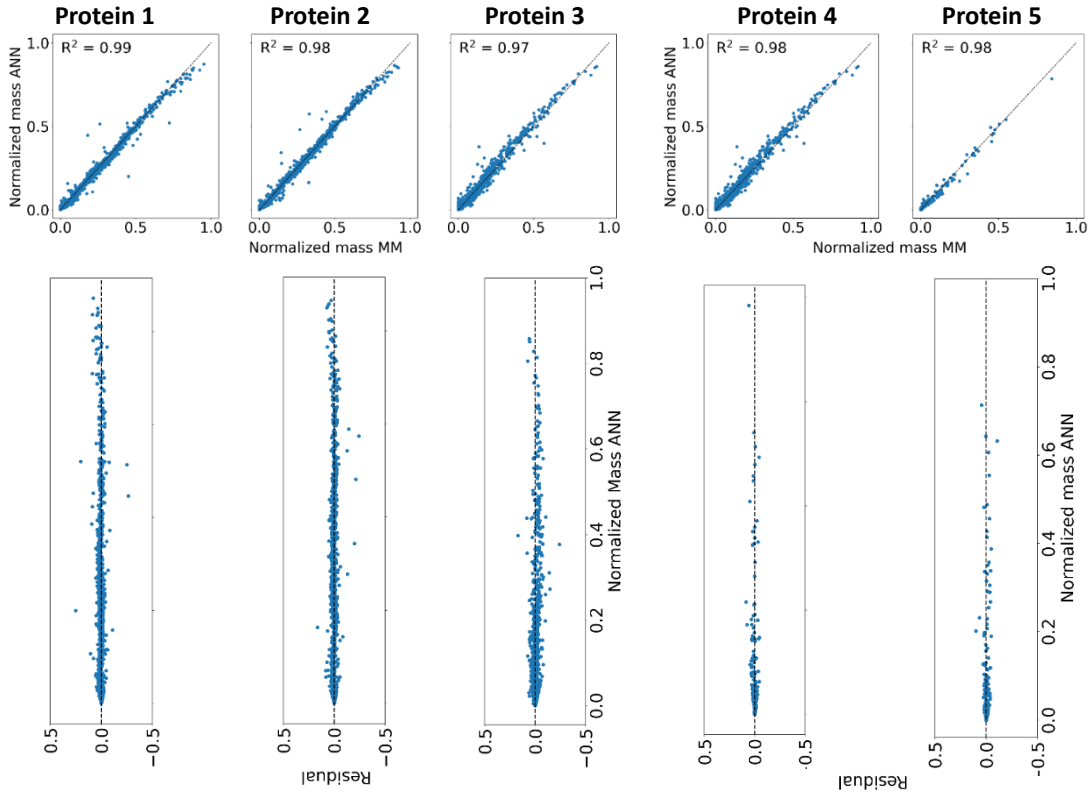


Figure 3.D.2. AEX mode, upper figure: Prediction capabilities for the normalized ANN outcome of mass against the outcome of MM. Lower figure: Residuals showing the difference between predicted mass values by the ANN and the MM. Both plots show unseen test-data (1493 points) for the proteins 1 to 5.

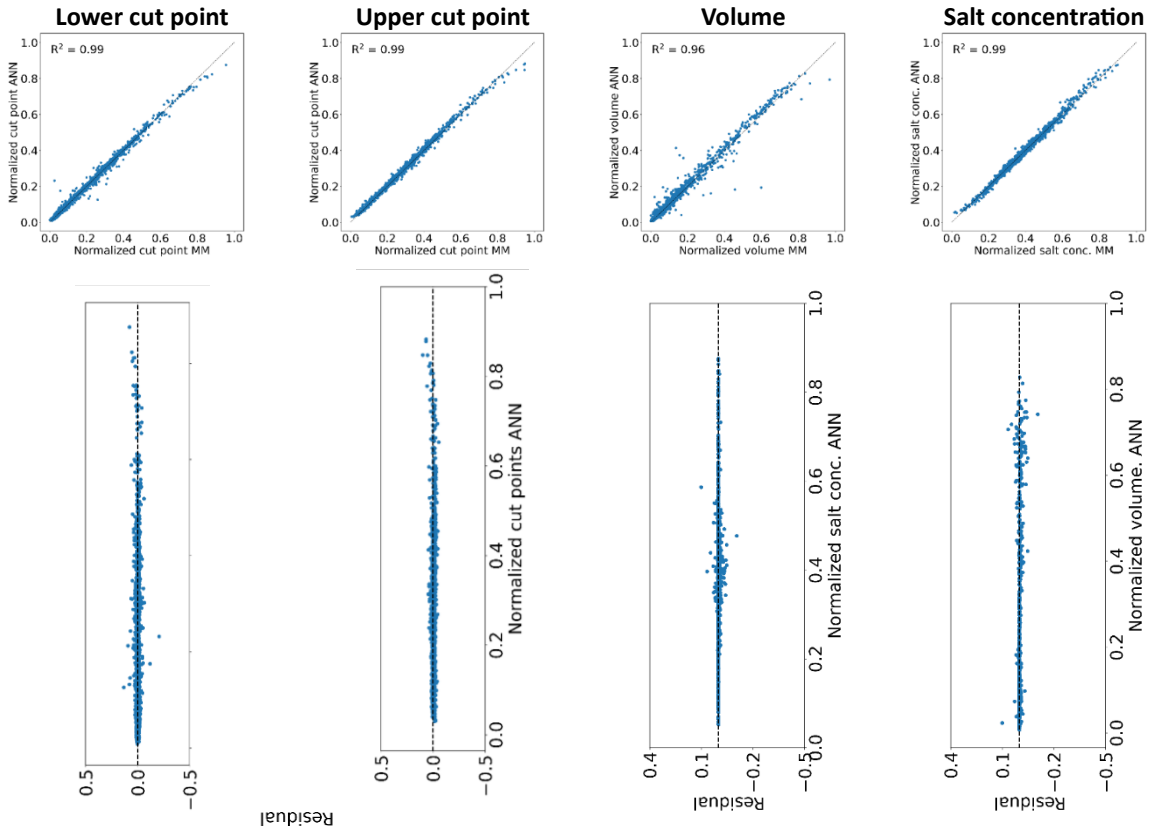


Figure 3.D.3. AEX mode, upper figure: Prediction capabilities for the normalized ANN outcome against the outcome of MM for the product pool volume and salt concentration, and both cut points. Lower figure: Residuals showing the difference between predicted values by the ANN and the MM. Both plots show unseen test-data (1493 points) for the product pool volume and salt concentration, and both cut points.

HIC

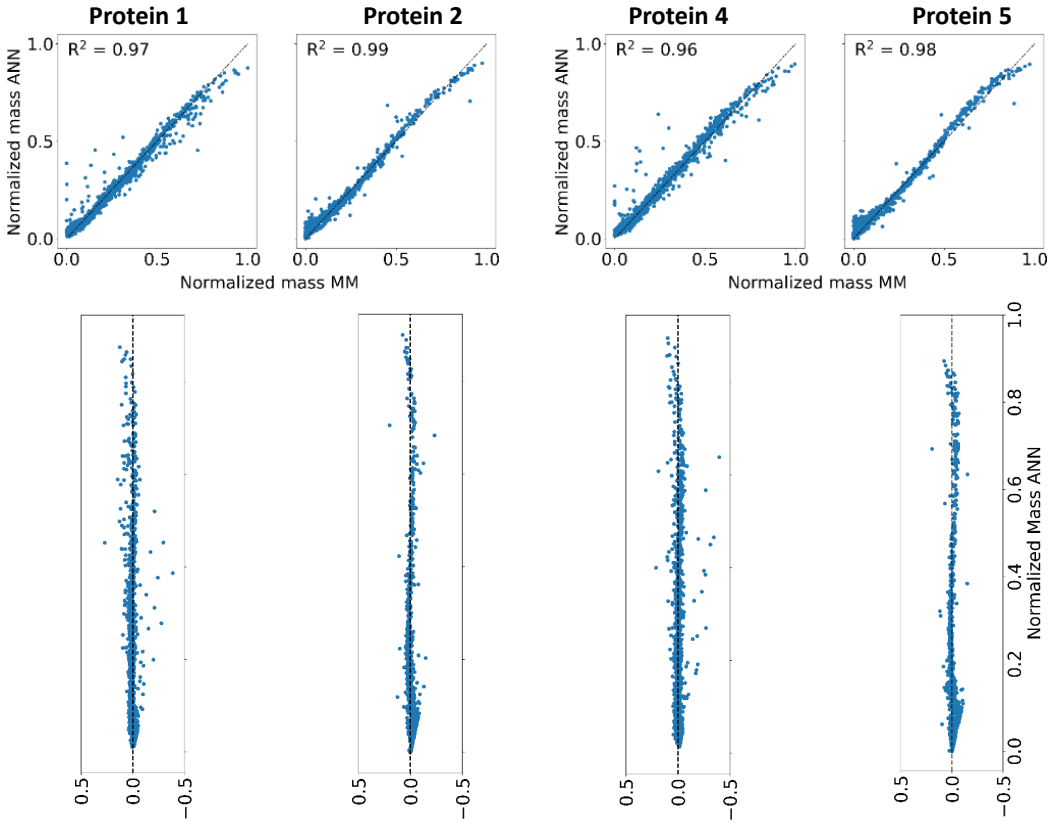


Figure 3.D.4. HIC mode, upper figure: Prediction capabilities for the normalized ANN outcome of mass against the outcome of MM. Lower figure: Residuals showing the difference between predicted mass values by the ANN and the MM. Both plots show unseen test-data (1462 points) for the proteins 1, 2, 4, and 5 for the HIC mode.

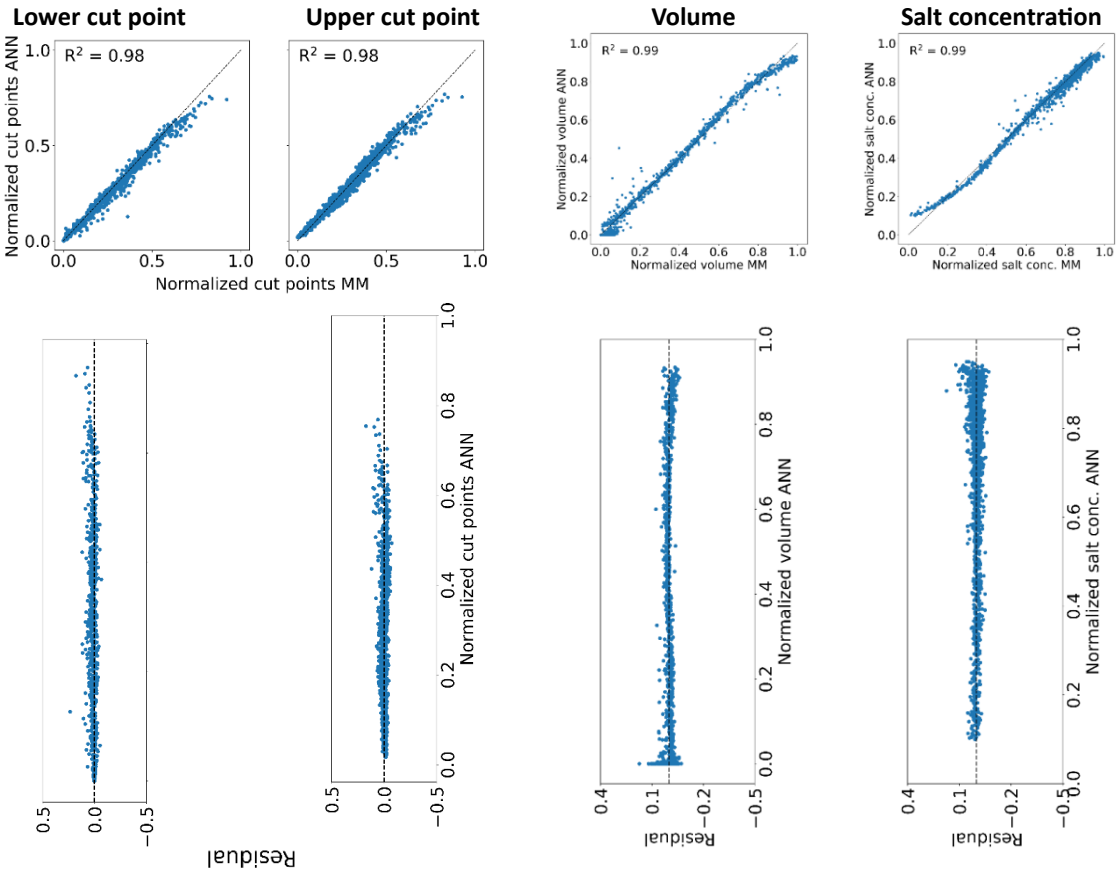


Figure 3.D.5. HIC mode, upper figure: Prediction capabilities for the normalized ANN outcome against the outcome of MM for the product pool volume and salt concentration, and both cut points. Lower figure: Residuals showing the difference between predicted values by the ANN and the MM. Both plots show unseen test-data (1493 points) for the product pool volume and salt concentration, and both cut points.

Appendix 3.E

Function evaluations to assess if the plateau has been reached. Figure 3.E.1. shows the function evaluations for the sequence of AEX – HIC for both the MM and the NN. Figure 3.E.2. shows the function evaluations of the ANN for the sequence of three unit operations.

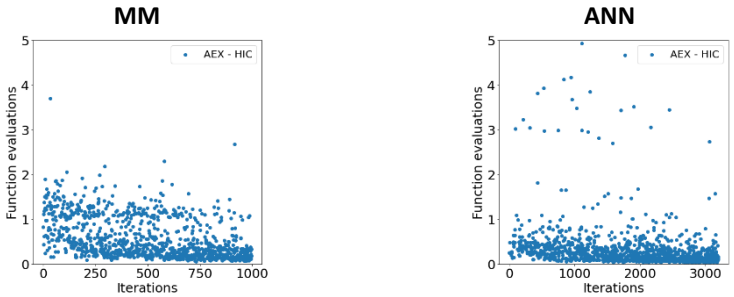


Figure 3.E.1. Function evaluations of the global optimization against the number of iterations for the MM (left) and ANN (right).

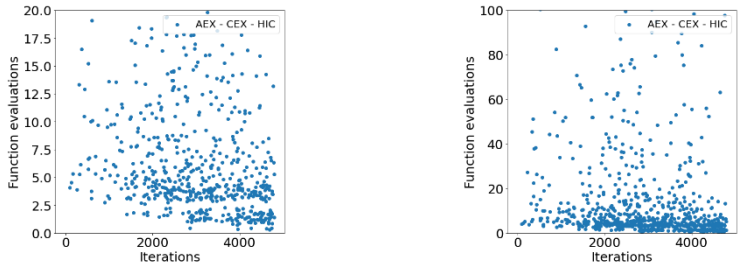


Figure 3.E.2. ANN function evaluations for the sequence of AEX – CEX – HIC, left showing the y-scale between 0 and 20 and right between 0 and 100.

Appendix - Chapter 4

Appendix 4.A

Details of the isotherm and resin parameters for each chromatography mode and all proteins are given in Table 4.A.1. The bed porosity was assumed to be 0.27 and the total porosity 0.95.

Table 4.A.1. Details of the isotherm and resin parameters used for the chromatography model [1].

	CEX	**AEX**	**HIC**
Resin	Mono S	Mono Q	Source PHE
Particle diameter (µm)	30	30	15
Pore diameter (nm)	40	40	83.4
Ligand density (Λ)	135	320	40

HIC column gradient experiments

Additional column gradient experiments (5, 10, and 15 CV) were performed on a HIC resin (PhenylFF) at pH = 7.0. The initial BSA concentration was 10 mg/mL using an injection loop of 500 µL and a flowrate of 1 mL/min. The initial buffer was a 3 M NaCl with 20 mM sodium phosphate buffer going to the final buffer of MilliQ. The experimental results are shown in Figure 4.A.1. BSA eluted in two peaks, one during the gradient and one after the gradient. The peak during the gradient was more critical as it is closer to the elution of the product of interest. As protein 3 (Chitotriosidase) also elutes at the end of the gradient, for simplicity, we assumed the same isotherm parameters for BSA.

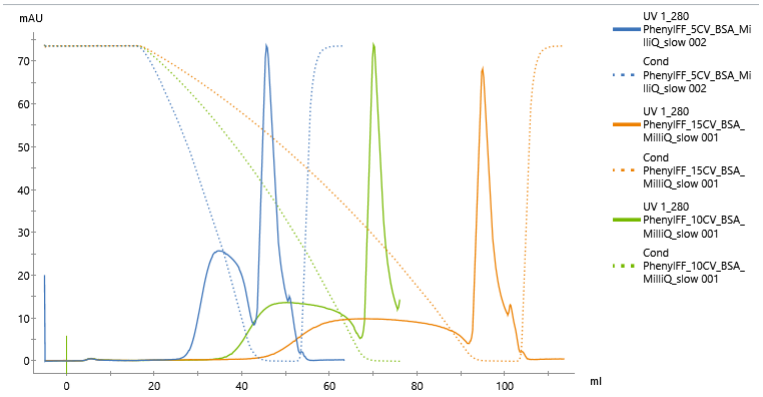


Figure 4.A.1. Experimental chromatograms of BSA on PhenylFF resin (HIC) eluting during the gradient lengths of 5 CV (blue), 10 CV (green), and 15 CV (orange). The dashed lines indicate the buffer and the solid lines the BSA solution in mAU.

[1] B.K. Nfor, T. Ahamed, M.W.H. Pinkse, L.A.M. van der Wielen, P.D.E.M. Verhaert, G.W.K. van Dedem, M.H.M. Eppink, E.J.A.X. van de Sandt, M. Ottens, 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) (2012) 3070-3083. https://doi.org/https://doi.org/10.1002/bit.24576.

Appendix 4.B

The transmembrane pressure is defined as

$$\Delta P_{TM} = \frac{P_{feed} + P_{retentate}}{2} - P_{permeate}\,. \qquad \textit{Eq. 4.B.1}$$

The osmotic pressure, π, in non-ideal solution is as follows:

$$\pi = R * T * C_{i,w} * \left(\frac{1}{M} + B_{22} * C_{i,w} \ldots\right), \qquad \textit{Eq. 4.B.2}$$

where R is the gas constant, T the temperature, $C_{i,w}$ is the solute wall concentration, and M is the molecular weight of the solute. The second virial coefficient, B_{22}, is generally sufficient to describe the osmotic pressure for low protein concentrations present in biopharmaceutical processes [3]. B_{22} values ($mL * mol/g^2$) were fitted to a second order polynomial function as function of the pH and salt concentration as

$$B_{22} = a_1 + a_2 * pH + a_3 * C_s + a_4 * pH * C_s + a_5 * pH^2 + a_6 * {C_s}^2. \qquad \textit{Eq. 4.B.3}$$

Data from Ma et al. [4] for Bovine Serum Albumin (BSA) under various pH and NaCl strengths was used to fit the constants; $a_1 = 6.801e^{-4}$, $a_2 = -2.215e^{-4}$, $a_3 = -9.696e^{-4}$, $a_4 = 1.075e^{-4}$, $a_5 = 1.913e^{-4}$, $a_6 = 1.804e^{-3}$.

The Sherwood number is used to determine the initial mass transfer coefficient, k_0, as

$$Sh = \frac{k_0 * d_h}{D} = aRe^b Sc^c \left(\frac{d_h}{l}\right)^d, \qquad \textit{Eq. 4.B.4}$$

where the Reynolds number is defined as $Re = \rho v d_h/u$, the Schmidt number as $Sc = \mu/\rho D$, in which ρ denotes the density, v is the cross-membrane velocity, and d_h is the hydraulic diameter. The diffusion coefficient, D, is determined by the Young correlation for global proteins [5]. The cross-membrane velocity, depending on the specific geometry of the system, is defined as $v = Q/(a_c \varepsilon_s)$, where a_c is the ratio of the feed channel area to membrane area and ε_s is the porosity of the spacer. In this work Screen type C was used for which (ε_s = 0.63, a_c = 0.0018)[6]. The constants, a, b, c, and d, of the Sherwood number relation (Eq. B.4) are empirical and determined based on the system configuration, in this work a rectangular channel with spacer ($a = 0.664$, $b = 1/2$, $c = 1/3$, and $d = 1/2$) [7]. The hydraulic diameter, d_h, is also system geometry dependent and defined as

$$d_h = 4h\,\frac{\varepsilon_s}{1 + \frac{2(1-\varepsilon_s)h}{r}}, \qquad \textit{Eq. 4.B.5}$$

where h is the half-height of the channel and r is the fibre radius, for a Screen type C spacer the values are $h = 0.026\ cm$ and $r = 0.014\ cm$ [6].

The solution viscosity is calculated through the Mooney's equation, which is often used for biophysical purposes, and given as follows [8, 9]:

$$\mu = \mu_0 \, e^{\left(\frac{[\mu]\, C_i}{1-\frac{C_i}{C_{max}}}\right)}, \qquad \textit{Eq. 4.B.6}$$

where $[\mu]$ is the intrinsic viscosity at low volume fractions, estimated at $8{\cdot}10^{-3}$ m^3/kg, and c_{max} is the maximum protein concentration, assumed to be 600 kg/m^3 [8]. A relation for the mass transfer coefficient as function of the protein concentration results from combining Eq. 18 and Eq. B.6 as

$$k = k_0 \, e^{\left(-\frac{1}{6}\frac{[\mu]\, C_p}{1-\frac{C_p}{C_{max}}}\right)}. \qquad \textit{Eq. 4.B.7}$$

Filtration experiments

The filtration experiment was performed as tangential flow filtration applying an ultrafiltration with variable volume diafiltration (UFVVD) mode [10]. An 88 cm^2 Millipore Pellicon 3 Ultracel 10 kDa membrane cassette was used, inside a membrane cassette holder (Merck Millipore). The schematic experimental set-up is shown in Figure 4.B.1. An Äkta Pure 25 system coupled with the Unicorn 7.0 software (Cytiva Life Sciences) was used to continuously monitor and collect data of the pH, conductivity, UV, and pressure before and after the membrane. The feed solution was pumped into the membrane unit, where the proteins were retained by the membrane and recycled back to the feed tank, while the water and salts could permeate. To perform the UF and or DF experiment, the diluent buffer was added to the feed tank using an additional HPLC pump (Shimadzu, UFLC/LC-20AD). Before each experiment, the retentate was recirculated for 15 minutes with a closed permeate stream to create a stable polarization layer. Scales were placed under the feed tank and permeate tank to confirm the mass balance of the volumes (Mettler Toledo).

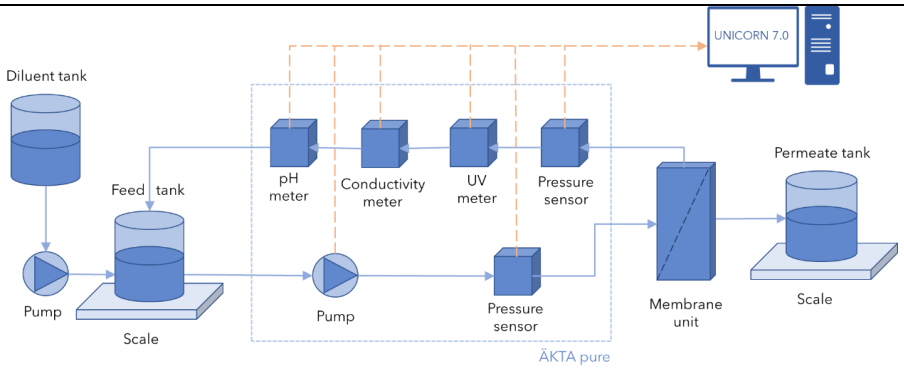


Figure 4.B.1. Schematic representation of the experimental UF/DF set-up using the Äkta Pure 25 system with Unicorn 7.0 for measuring the pH, conductivity, UV, and pressure. The blue lines indicate the flow streams, while the dashed yellow lines are the data streams.

The filtration model was validated for an UF/DF operation with a 0.3 mg/mL BSA solution (Sigma Aldrich) exchanging from sodium phosphate buffer (10 mM, pH = 7), containing 175 mM NaCl to a 0 mM NaCl potassium phosphate buffer (10 mM, pH = 6.5). The flowrate was 20 mL/min, as this was the maximum flowrate possible in an Äkta Pure 25 system. UV measurements were conducted at a wavelength of 280 nm. All buffer solutions were filtered with a MediaKap 0.2 µm pore-size hollow fibre media filter to remove undissolved salts and afterwards degassed using a Branson ultrasonic bath.

Membrane resistance

Through water flux experiments at a flowrate of 20 mL/min, the water permeate flux, J_w, was determined at $19.7 \cdot 10^{-6}$ m/s. Subsequently, the initial membrane resistance was calculated to be $7.6 \cdot 10^{12}\ m^{-1}$ using the following formula:

$$R_m = \frac{\Delta P}{\mu_w * J_w}, \qquad \textit{Eq. 4.B.8}$$

where μ_w is the viscosity of water, and the applied transmembrane pressure, ΔP, was 0.15 MPa.

[3] R. van Reis, E.M. Goodrich, C.L. Yson, L.N. Frautschy, R. Whiteley, A.L. Zydney, Constant Cwall ultrafiltration process control, J Membrane Sci 130(1) (1997) 123-140. https://doi.org/https://doi.org/10.1016/S0376-7388(97)00012-4.

[4] Y. Ma, D.M. Acosta, J.R. Whitney, R. Podgornik, N.F. Steinmetz, R.H. French, V.A. Parsegian, Determination of the second virial coefficient of bovine serum albumin under varying pH and ionic strength by composition-gradient multi-angle static light scattering, Journal of Biological Physics 41(1) (2015) 85-97. https://doi.org/10.1007/s10867-014-9367-7.

[5] M.E. Young, P.A. Carroad, R.L. Bell, Estimation of diffusion coefficients of proteins, Biotechnology and Bioengineering 22(5) (1980) 947-955. https://doi.org/10.1002/bit.260220504.

[6] H. Lutz, 3 - Modules, in: H. Lutz (Ed.), Ultrafiltration for Bioprocessing, Woodhead Publishing, Oxford, 2015, pp. 31-43. https://doi.org/https://doi.org/10.1016/B978-1-907568-46-6.00003-3.

[7] L.J. Zeman, A.L. Zydney, Microfiltration and Ultrafiltration: Principles and Applications, 1996. https://doi.org/https://doi.org/10.1201/9780203747223.

[8] M. Zidar, P. Rozman, K. Belko-Parkel, M. Ravnik, Control of viscosity in biopharmaceutical protein formulations, Journal of Colloid and Interface Science 580 (2020) 308-317. https://doi.org/https://doi.org/10.1016/j.jcis.2020.06.105.

[9] M. Mooney, The viscosity of a concentrated suspension of spherical particles, Journal of Colloid Science 6(2) (1951) 162-170. https://doi.org/https://doi.org/10.1016/0095-8522(51)90036-0.

[10] G.A. Foley, Membrane Filtration: A Problem Solving Approach with MATLAB, 2013.

Appendix 4.C

Optimized flowsheet results from the global optimization for each optimization strategy (e.g., simultaneous, top-to-bottom, decomposition) using MMs or ANNs are shown in Table 4.C.1. – 4.C.6. Each element in a flowsheet, second column in each table, is an integer number between 1 to 5, which in this study represents the considered unit operations; CEX, AEX, HIC, dilution, and filtration.

Table 4.C.1. Global optimization results of the simultaneous strategy using MMs. Host cell proteins are indicated as HCP.

Simultaneous		Purity (%)	Yield (%)	HCP level (ng/$mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
0	[1]	34.54	92.96	1895250.0	64.76	10.18	1.28	65
1	[1-4-2]	91.03	89.13	98598.90	98.24	2.85	5.65	91
2	[1-4-2-4-3]	98.01	95.14	20316.00	99.61	11.33	10.92	95
3	[1-4-2-5-3]	98.92	92.10	10967.00	99.80	8.81	9.02	95
4	[1-4-3]	79.78	75.11	253482.00	96.19	8.88	5.81	81
5	[1-4-3-4-2]	99.65	90.28	3523.37	99.94	3.82	8.70	95
6	[1-4-3-5-2]	99.51	93.68	4945.04	99.91	3.36	11.43	96
7	[1-5-2]	90.37	84.66	106550.00	98.20	7.48	3.14	90
8	[1-5-2-4-3]	99.87	83.13	1326.59	99.98	7.72	7.17	93
9	[1-5-2-5-3]	99.79	87.74	2140.07	99.96	6.52	11.01	94
10	[1-5-3]	81.20	68.42	231556.00	96.83	8.59	4.55	80
11	[1-5-3-4-2]	99.66	95.98	3366.35	99.94	8.71	8.59	97
12	[1-5-3-5-2]	98.57	92.17	14477.10	99.73	4.72	8.47	95
13	[2]	41.54	99.84	1407350.0	71.90	1.66	3.25	70
14	[2-4-1]	91.64	85.45	91218.90	98.44	8.19	7.31	90
15	[2-4-1-4-3]	99.44	85.44	5611.53	99.90	13.04	6.10	94
16	[2-4-1-5-3]	98.30	95.55	17261.00	99.67	5.60	8.43	96
17	[2-4-3]	62.20	100.00	600092.00	88.00	7.70	5.99	80
18	[2-4-3-4-1]	99.86	85.73	1382.80	99.98	6.08	6.61	94
19	[2-4-3-5-1]	99.71	79.26	2882.19	99.95	5.69	7.28	92
20	[2-5-1]	91.10	81.89	97718.50	98.40	8.67	3.34	89
21	[2-5-1-4-3]	98.62	80.29	14033.90	99.77	10.43	4.35	93
22	[2-5-1-5-3]	99.66	91.74	3443.40	99.94	11.68	8.11	96
23	[2-5-3]	62.02	100.00	600235.00	88.00	9.78	3.11	80
24	[2-5-3-4-1]	99.86	89.30	1391.22	99.98	5.72	6.48	95
25	[2-5-3-5-1]	98.82	87.84	11990.80	99.79	4.80	8.25	94
26	[3]	29.97	100.50	2336510.0	53.04	10.76	1.26	65
27	[3-4-1]	78.11	81.13	280324.00	95.45	13.67	2.51	83
28	[3-4-1-4-2]	99.70	74.89	3014.04	99.95	9.15	6.75	91
29	[3-4-1-5-2]	97.91	89.16	21305.60	99.62	3.79	5.10	95
30	[3-4-2]	61.90	96.60	615597.00	88.11	4.50	4.35	79
31	[3-4-2-4-1]	99.68	87.65	3203.55	99.94	3.17	7.70	95
32	[3-4-2-5-1]	99.82	80.81	1810.59	99.97	8.25	6.33	93
33	[3-5-1]	79.74	73.97	254082.00	96.24	10.58	5.02	81

Simultaneous		Purity (%)	Yield (%)	HCP level (ng/mg$_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/g$_{product}$)	WOP
34	[3-5-1-4-2]	99.51	82.64	4970.18	99.92	6.70	8.31	93
35	[3-5-1-5-2]	99.44	82.47	5608.53	99.91	6.43	5.08	93
36	[3-5-2]	61.88	90.95	616000.00	88.80	4.49	6.05	77
37	[3-5-2-4-1]	99.39	86.45	6168.08	99.89	1.88	9.60	94
38	[3-5-2-5-1]	99.36	73.75	6458.24	99.90	5.23	6.71	90

Table 4.C.2. Global optimization results of the top-to-bottom strategy using MMs. Host cell proteins are indicated as HCP.

Top-to-bottom		Purity	Yield	HCP level (ng/mg$_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/g$_{product}$)	WOP
0	[1]	34.83	92.68	1871500	65.31	10.59	1.84	65
1	[1-4-2]	86.60	92.45	154736.0	97.14	4.32	4.41	90
2	[1-4-2-4-3]	92.73	93.81	78444.90	98.53	9.58	5.39	93
3	[1-4-2-5-3]	92.73	93.81	78444.90	98.53	9.58	5.39	93
4	[1-4-3]	69.55	92.95	437905.0	91.86	9.39	2.40	82
5	[1-4-3-4-2]	92.71	92.27	78677.70	98.55	10.63	7.63	93
6	[1-4-3-5-2]	92.70	87.66	78691.40	98.62	10.00	3.49	92
7	[1-5-2]	86.60	87.85	154714.0	97.28	4.07	4.00	89
8	[1-5-2-4-3]	92.73	92.75	78350.20	98.55	9.09	4.41	93
9	[1-5-2-5-3]	92.73	92.75	78350.20	98.55	9.09	4.41	93
10	[1-5-3]	69.54	88.27	437946.0	92.27	9.11	2.53	81
11	[1-5-3-4-2]	92.73	88.14	78372.90	98.62	3.35	3.82	92
12	[1-5-3-5-2]	92.73	83.72	78385.40	98.69	3.20	3.87	91
13	[2]	41.32	98.89	1419880	71.92	1.58	2.31	70
14	[2-4-1]	90.77	89.36	101701.0	98.18	10.52	19.06	88
15	[2-4-1-4-3]	98.82	89.50	11981.60	99.79	9.08	20.26	92
16	[2-4-1-5-3]	98.82	85.05	11980.50	99.80	8.68	21.36	91
17	[2-4-3]	62.37	100.00	600007.0	88.00	10.67	3.47	80
18	[2-4-3-4-1]	99.28	95.69	7224.94	99.86	3.50	5.77	97
19	[2-4-3-5-1]	99.31	90.86	6910.50	99.87	3.32	5.86	96
20	[2-5-1]	90.95	85.01	99513.40	98.31	9.99	5.44	90
21	[2-5-1-4-3]	98.74	84.95	12755.50	99.78	8.51	6.72	94
22	[2-5-1-5-3]	98.74	84.95	12755.50	99.78	8.51	6.72	94
23	[2-5-3]	62.37	100.00	600007.0	88.00	10.67	3.47	80
24	[2-5-3-4-1]	99.20	95.70	8056.44	99.85	3.57	5.78	97
25	[2-5-3-5-1]	99.26	90.77	7461.59	99.86	3.44	5.94	96
26	[3]	29.91	98.81	2343730	53.69	11.19	1.38	64
27	[3-4-1]	78.48	80.89	274155.0	95.56	11.49	3.40	83
28	[3-4-1-4-2]	97.85	80.51	22024.90	99.65	6.51	8.57	91
29	[3-4-1-5-2]	97.84	76.50	22037.90	99.66	6.19	4.77	91
30	[3-4-2]	62.32	98.49	604715.0	88.09	5.62	3.60	80

Top-to-bottom		Purity	Yield	HCP level (ng/$mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/$g_{product}$)	WOP
31	[3-4-2-4-1]	98.08	94.45	19562.70	99.63	2.23	5.69	96
32	[3-4-2-5-1]	99.18	89.33	8234.36	99.85	2.46	5.70	95
33	[3-5-1]	78.96	76.99	266473.0	95.90	10.91	3.12	82
34	[3-5-1-4-2]	97.70	76.10	23491.00	99.64	8.78	4.62	91
35	[3-5-1-5-2]	97.71	72.36	23486.50	99.66	8.34	4.12	90
36	[3-5-2]	62.32	93.57	604630.0	88.68	5.34	2.81	79
37	[3-5-2-4-1]	98.74	89.43	12740.60	99.77	3.57	5.45	95
38	[3-5-2-5-1]	99.07	84.98	9357.07	99.84	3.62	5.12	94

Table 4.C.3. Global optimization results of the decomposition strategy using MMs. Host cell proteins are indicated as HCP.

Decomposition		Purity	Yield	HCP level (ng/$mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/$g_{product}$)	WOP
0	[1]	34.52	93.28	1897250	64.61	10.58	1.59	65
1	[1-4-2]	92.08	85.81	86032.80	98.52	3.11	3.99	91
2	[1-4-2-4-3]	98.11	92.60	19232.10	99.64	9.41	4.89	96
3	[1-4-2-5-3]	98.11	92.60	19232.10	99.64	9.41	4.89	96
4	[1-4-3]	80.28	76.69	245586.0	96.23	11.32	3.34	82
5	[1-4-3-4-2]	97.46	76.13	26044.50	99.60	4.14	5.11	91
6	[1-4-3-5-2]	97.46	72.33	26046.10	99.62	3.92	5.03	89
7	[1-5-2]	92.40	81.29	82251.50	98.66	2.44	5.25	90
8	[1-5-2-4-3]	99.05	88.93	9637.28	99.83	9.20	6.00	95
9	[1-5-2-5-3]	99.05	88.93	9637.28	99.83	9.20	6.00	95
10	[1-5-3]	77.91	81.54	283451.0	95.38	8.64	2.85	83
11	[1-5-3-4-2]	99.22	81.40	7909.04	99.87	4.37	5.02	93
12	[1-5-3-5-2]	99.22	77.32	7902.60	99.88	4.17	4.73	92
13	[2]	41.49	99.56	1410340	71.92	1.83	3.05	70
14	[2-4-1]	90.09	89.03	110042.0	98.04	9.87	7.15	90
15	[2-4-1-4-3]	98.30	89.78	17334.00	99.69	9.05	8.24	94
16	[2-4-1-5-3]	98.30	89.78	17334.00	99.69	9.05	8.24	94
17	[2-4-3]	61.86	100.00	600000.0	88.00	9.86	4.14	80
18	[2-4-3-4-1]	99.07	96.17	9345.88	99.82	2.87	6.04	97
19	[2-4-3-5-1]	99.19	91.26	8184.55	99.85	2.81	6.22	96
20	[2-5-1]	93.33	78.97	71420.20	98.87	11.65	3.29	90
21	[2-5-1-4-3]	99.32	79.18	6882.03	99.89	8.40	3.99	93
22	[2-5-1-5-3]	99.32	75.24	6881.85	99.90	7.88	4.21	91
23	[2-5-3]	62.38	100.00	600000.0	88.00	11.44	3.49	80
24	[2-5-3-4-1]	99.08	96.18	9307.90	99.82	2.09	5.72	97
25	[2-5-3-5-1]	99.12	91.29	8848.74	99.84	2.00	5.94	96
26	[3]	29.94	99.71	2339510	53.34	10.45	1.60	65
27	[3-4-1]	78.66	80.00	271281.0	95.66	12.44	3.49	83
28	[3-4-1-4-2]	98.02	79.27	20160.10	99.68	5.39	5.23	92

	Decomposition	Purity	Yield	HCP level ($ng/mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
29	[3-4-1-5-2]	98.02	75.29	20179.60	99.70	5.11	4.80	91
30	[3-4-2]	62.10	94.14	610333.0	88.51	10.06	2.71	79
31	[3-4-2-4-1]	98.47	90.70	15512.70	99.72	3.03	5.16	95
32	[3-4-2-5-1]	99.10	85.83	9109.09	99.84	3.17	5.26	94
33	[3-5-1]	81.53	70.48	226555.0	96.81	10.19	4.04	81
34	[3-5-1-4-2]	98.31	70.33	17198.30	99.76	3.18	5.42	89
35	[3-5-1-5-2]	98.31	66.81	17190.10	99.77	3.10	5.48	88
36	[3-5-2]	62.38	92.44	603106.0	88.85	3.41	3.86	78
37	[3-5-2-4-1]	98.69	87.63	13321.80	99.77	3.21	6.43	94
38	[3-5-2-5-1]	99.24	83.12	7625.05	99.87	3.36	6.42	93

Table 4.C.4. Global optimization results of the simultaneous strategy using ANNs. Host cell proteins are indicated as HCP.

	Simultaneous	Purity	Yield	HCP level ($ng/mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
0	[1]	37.75	100.00	1649240	67.02	23.52	1.03	69
1	[1-4-2]	87.66	98.48	140750.0	97.23	13.72	3.01	93
2	[1-4-2-4-3]	99.60	100.00	4046.83	99.92	14.08	6.90	98
3	[1-4-2-5-3]	99.36	100.00	6452.30	99.87	18.74	7.74	98
4	[1-4-3]	71.19	100.00	404615.0	91.91	19.64	4.96	85
5	[1-4-3-4-2]	98.87	98.10	11383.50	99.78	9.38	7.29	97
6	[1-4-3-5-2]	99.36	98.59	6429.60	99.87	6.12	8.43	98
7	[1-5-2]	87.35	93.54	144796.0	97.29	8.76	4.88	91
8	[1-5-2-4-3]	99.72	97.17	2798.96	99.95	16.23	9.43	97
9	[1-5-2-5-3]	99.23	91.03	7778.06	99.86	11.04	8.98	95
10	[1-5-3]	82.58	77.66	210956.0	96.72	10.60	5.52	83
11	[1-5-3-4-2]	95.86	100.00	43145.90	99.14	10.51	7.01	97
12	[1-5-3-5-2]	99.08	89.62	9316.75	99.83	4.87	9.53	95
13	[2]	46.29	100.00	1160110	76.80	16.28	1.02	73
14	[2-4-1]	90.95	100.00	99540.70	98.01	12.71	6.09	94
15	[2-4-1-4-3]	99.43	99.14	5755.11	99.89	17.31	8.04	98
16	[2-4-1-5-3]	99.56	100.00	4387.61	99.91	15.83	8.52	98
17	[2-4-3]	80.25	100.00	246093.0	95.08	19.27	7.27	89
18	[2-4-3-4-1]	99.64	100.00	3606.33	99.93	11.68	7.23	98
19	[2-4-3-5-1]	99.53	89.57	4753.26	99.91	3.89	8.82	95
20	[2-5-1]	84.92	87.78	177549.0	96.88	6.60	6.09	88
21	[2-5-1-4-3]	99.60	96.46	3988.93	99.92	11.99	10.59	97
22	[2-5-1-5-3]	99.74	83.78	2610.75	99.96	12.32	10.30	93
23	[2-5-3]	78.78	98.93	269351.0	94.67	18.74	7.39	88
24	[2-5-3-4-1]	99.52	100.00	4872.64	99.90	15.23	7.82	98
25	[2-5-3-5-1]	98.28	93.96	17483.60	99.67	3.84	9.13	96

Simultaneous		Purity	Yield	HCP level ($ng/mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
26	[3]	39.65	100.00	1521850	69.56	17.74	1.42	70
27	[3-4-1]	81.43	99.16	228056.0	95.48	12.05	4.58	90
28	[3-4-1-4-2]	99.69	100.00	3114.43	99.94	7.56	6.96	98
29	[3-4-1-5-2]	98.85	94.99	11653.70	99.78	6.45	7.22	96
30	[3-4-2]	77.52	100.00	289954.0	94.20	7.88	3.20	88
31	[3-4-2-4-1]	99.60	100.00	3991.65	99.92	6.69	7.60	98
32	[3-4-2-5-1]	97.45	87.52	26177.20	99.54	3.55	6.34	94
33	[3-5-1]	84.82	64.64	178914.0	97.69	10.31	6.17	81
34	[3-5-1-4-2]	98.62	84.44	13946.50	99.76	9.17	7.53	93
35	[3-5-1-5-2]	99.01	83.53	9955.00	99.83	3.01	9.66	93
36	[3-5-2]	73.57	91.65	359284.0	93.41	5.35	3.59	84
37	[3-5-2-4-1]	99.19	88.97	8196.73	99.85	4.58	8.52	95
38	[3-5-2-5-1]	96.93	78.81	31639.60	99.50	2.59	8.61	90

Table 4.C.5. Global optimization results of the top-to-bottom strategy using ANNs. Host cell proteins are indicated as HCP. The flowsheet [2-4-1] could not be optimized as the dilution step after the first unit operation caused an out-of-range parameter value for the loading factor of the subsequent unit operation, therefore Not-a-number (Nan) appeared. The flowsheets [2-4-1-4-3] and [2-4-1-5-3] depend on the outcome of the previous result from [2-4-1], as no solution was found, these two flowsheets could not be optimized either.

Top-to-bottom		Purity	Yield	HCP level ($ng/mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
0	[1]	36.70	100.00	1724570	65.51	23.44	1.05	68
1	[1-4-2]	81.97	100.00	219984.0	95.60	16.96	2.10	91
2	[1-4-2-4-3]	89.68	100.00	115034.0	97.70	13.68	3.94	94
3	[1-4-2-5-3]	89.57	93.94	116386.0	97.81	12.91	4.20	92
4	[1-4-3]	70.10	100.00	426636.0	91.47	12.51	3.26	84
5	[1-4-3-4-2]	89.65	100.00	115500.0	97.69	13.03	4.20	94
6	[1-4-3-5-2]	89.41	92.73	118398.0	97.80	12.51	4.45	92
7	[1-5-2]	81.60	92.69	225513.0	95.82	17.06	2.05	88
8	[1-5-2-4-3]	89.45	92.69	117945.0	97.81	11.72	4.13	92
9	[1-5-2-5-3]	89.45	88.06	117998.0	97.92	11.17	4.35	90
10	[1-5-3]	70.00	94.56	428668.0	91.89	11.84	3.43	83
11	[1-5-3-4-2]	89.60	94.56	116118.0	97.80	11.06	4.72	92
12	[1-5-3-5-2]	89.25	86.66	120451.0	97.91	10.65	5.02	90
13	[2]	46.80	100.00	1136730	77.27	21.09	1.11	73
14	[2-4-1]	Nan	Nan	Nan	Nan	Nan	Nan	0
15	[2-4-1-4-3]	Nan	Nan	Nan	Nan	Nan	Nan	0
16	[2-4-1-5-3]	Nan	Nan	Nan	Nan	Nan	Nan	0
17	[2-4-3]	77.22	100.00	294985.0	94.10	13.26	4.50	88
18	[2-4-3-4-1]	98.33	100.00	17002.50	99.66	2.57	8.70	97
19	[2-4-3-5-1]	98.64	92.90	13835.10	99.74	2.44	7.87	96

Top-to-bottom		Purity	Yield	HCP level (ng/$mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/$g_{product}$)	WOP
20	[2-5-1]	86.09	85.80	161565.0	97.23	21.87	2.07	88
21	[2-5-1-4-3]	99.05	85.80	9546.57	99.84	11.30	4.51	94
22	[2-5-1-5-3]	99.05	81.51	9603.52	99.84	10.77	4.74	93
23	[2-5-3]	76.70	95.00	303709.0	94.23	12.61	4.72	86
24	[2-5-3-4-1]	Nan	Nan	Nan	Nan	Nan	Nan	0
25	[2-5-3-5-1]	98.09	88.66	19479.90	99.65	1.87	8.65	94
26	[3]	36.57	100.00	1734280	65.31	30.98	1.02	68
27	[3-4-1]	75.28	97.93	328322.0	93.57	23.04	1.71	87
28	[3-4-1-4-2]	97.02	97.93	30717.00	99.40	10.86	3.03	97
29	[3-4-1-5-2]	96.82	87.25	32837.50	99.43	10.39	3.17	94
30	[3-4-2]	73.69	100.00	357070.0	92.86	8.62	3.20	86
31	[3-4-2-4-1]	97.96	100.00	20830.20	99.58	2.58	6.97	98
32	[3-4-2-5-1]	98.51	92.73	15081.30	99.72	2.38	6.31	96
33	[3-5-1]	76.76	88.25	302690.0	94.66	21.04	1.95	84
34	[3-5-1-4-2]	97.30	88.25	27723.20	99.51	5.00	4.65	94
35	[3-5-1-5-2]	97.15	79.43	29349.10	99.53	4.70	4.98	91
36	[3-5-2]	72.92	92.81	371351.0	93.11	8.09	3.72	84
37	[3-5-2-4-1]	Nan	Nan	Nan	Nan	Nan	Nan	0
38	[3-5-2-5-1]	97.96	83.10	20827.00	99.65	4.69	7.21	92

Table 4.C.6. Global optimization results of the decomposition strategy using ANNs. Host cell proteins are indicated as HCP.

Decomposition		Purity	Yield	HCP level (ng/$mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption (L/$g_{product}$)	WOP
0	[1]	36.69	100.00	1725650.0	65.49	23.21	1.05	68
1	[1-4-2]	87.80	97.72	138981.00	97.28	10.74	3.29	93
2	[1-4-2-4-3]	97.47	100.00	25953.80	99.48	12.36	5.12	98
3	[1-4-2-5-3]	97.47	100.00	25953.80	99.48	12.36	5.12	98
4	[1-4-3]	70.58	92.08	416888.00	92.32	14.92	2.56	82
5	[1-4-3-4-2]	87.04	92.08	148886.00	97.26	11.10	3.90	90
6	[1-4-3-5-2]	86.59	84.13	154897.00	97.39	10.52	4.16	88
7	[1-5-2]	87.78	92.35	139224.00	97.43	6.58	4.14	91
8	[1-5-2-4-3]	96.92	98.01	31743.40	99.38	11.72	6.38	97
9	[1-5-2-5-3]	96.92	98.01	31743.40	99.38	11.72	6.38	97
10	[1-5-3]	68.01	95.36	470287.00	91.03	14.92	5.55	82
11	[1-5-3-4-2]	90.18	95.36	108947.00	97.92	13.28	6.70	92
12	[1-5-3-5-2]	89.63	85.42	115640.00	98.02	12.39	7.36	89
13	[2]	44.42	100.00	1251280.0	74.97	5.25	3.01	72
14	[2-4-1]	90.44	100.00	105713.00	97.89	14.23	5.70	94
15	[2-4-1-4-3]	99.22	100.00	7876.76	99.84	13.64	7.60	98
16	[2-4-1-5-3]	99.22	100.00	7876.76	99.84	13.64	7.60	98

	Decomposition	Purity	Yield	HCP level ($ng/mg_{product}$)	HCP clearance (%)	Product concentration (g/L)	Buffer consumption ($L/g_{product}$)	WOP
17	[2-4-3]	81.04	100.00	233934.00	95.32	18.75	7.52	89
18	[2-4-3-4-1]	97.34	100.00	27366.90	99.45	2.23	10.76	97
19	[2-4-3-5-1]	97.60	94.57	24610.30	99.53	2.08	11.01	95
20	[2-5-1]	87.85	85.28	138318.00	97.64	7.90	6.26	88
21	[2-5-1-4-3]	97.71	87.82	23443.10	99.59	10.21	8.34	94
22	[2-5-1-5-3]	97.71	87.82	23443.10	99.59	10.21	8.34	94
23	[2-5-3]	79.32	99.95	260652.00	94.79	20.19	6.54	88
24	[2-5-3-4-1]	97.25	99.95	28308.50	99.43	2.17	11.11	96
25	[2-5-3-5-1]	98.13	93.03	19052.10	99.65	1.96	10.54	95
26	[3]	42.62	99.75	1346510.0	73.14	23.19	1.37	71
27	[3-4-1]	84.11	98.45	188961.00	96.28	23.50	4.28	91
28	[3-4-1-4-2]	99.57	98.45	4289.14	99.92	7.91	6.04	98
29	[3-4-1-5-2]	99.55	90.79	4495.57	99.92	7.59	6.42	96
30	[3-4-2]	77.49	99.40	290569.00	94.22	12.07	4.60	88

Appendix - Chapter 5

Appendix 5.A

Dead volume and dwell volume

The volume of the tubing was determined by excluding the column and using 1 M sodium chloride with a 100 µL sample loop. A schematic overview of the tubing in the Äkta system is shown in Figure 5.A.1, in which the dead volume is indicated from the numbers 2 to 4 and the dwell volume from 1 to 3.

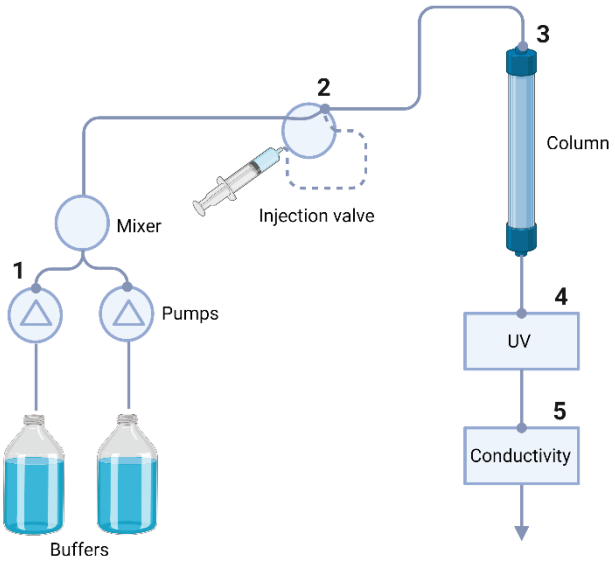


Figure 5.A.1. Schematic representation of the Äkta system, the dead volume is defined from point 2 to 4 and the dwell volume from point 1 to 3. The injection valve is indicated with the dashed line and not considered in the dead volume and dwell volume. Created with Biorender.com.

The dead volume (V_{dead}), tubing 3 and 4, is calculated according to Schmidt-Traub et al. (2012) as follows [1]:

$$V_{dead} = V_{R,0} - \frac{V_{inj}}{2} - V_5, \qquad \textit{Eq. 5.A.1}$$

where $V_{R,0}$ is the retention volume measured including the injection volume (V_{inj}), which is therefore subtracted to only obtain the dead volume. V_5 is the tubing between the UV-detector and the conductivity (indicated with number 5), from the internal diameter, 0.50 mm, and the length, 170 mm, it was calculated to be 0.033 mL.

The dwell volume is needed for the calculations in the regression formula and is equal to the volume from point 1 to 3 (Figure 5.A.1). The tubing before point 1 is already filled prior to elution. The dwell volume was determined by introducing buffer B, containing 1 M sodium chloride as a pulse for 5 CV, followed by subtracting the V_{dead} and V_5.

Porosity calculations

The total porosity (ε_t) was determined using 1 M sodium chloride, as salt can enter the pores, and calculated as follows:

$$\varepsilon_t = \frac{V_m + V_{pore}}{V_C} \qquad \textit{Eq. 5.A.2}$$

$$V_m + V_{pore} = V_{0,ret} - V_{dead} \qquad \textit{Eq. 5.A.3}$$

where V_m is the interstitial volume of the fluid phase also known as the column void volume, V_{pore} is the volume of the pore system, and V_C is the total volume of the packed column. $V_{0,ret}$ is the measured retention volume from which the dead volume is subtracted to only consider the retention volume in the column. The external porosity, $\varepsilon_b = V_m/V_C$, was determined using a solution of 10 mg/mL Dextran (DXT1740K, American Polymer Standards Corporation, USA) with a volume of 250 µL. V_m was determined using Eq. 5.A.3. Subsequently, the total and external porosity are used to determine the internal porosity (ε_p) as

$$\varepsilon_p = \frac{\varepsilon_t - \varepsilon_b}{1 - \varepsilon_b}. \qquad \textit{Eq. 5.A.4}$$

[1] H. Schmidt-Traub, M. Schulte, A. Seidel-Morgenstern, H. Schmidt-Traub, Preparative chromatography, Wiley Online Library2012.

Appendix 5.B

Regression plots of each protein at pH 3.5, 4.3, 5.0, and 7.0 corresponding to the Figures 5.B.1 – 5.B.4 respectively.

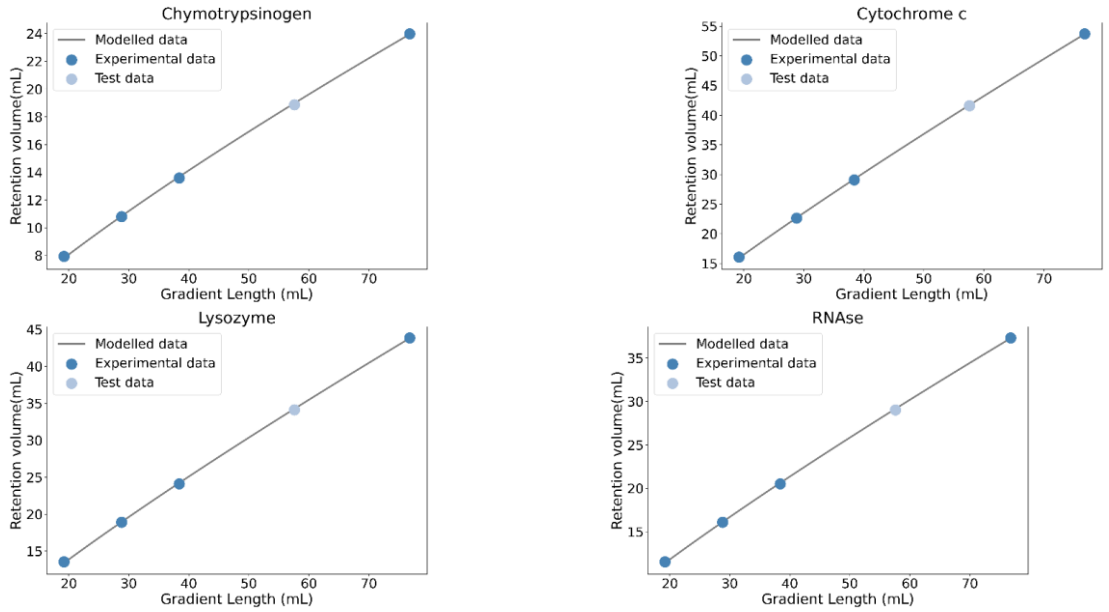


Figure 5.B.1. Fitted regression curves at pH 3.5 (grey line) of the experimental data (dark blue dots) and the test data point (ligth blue dot) at 58.2 mL, equal to 60 CV as 1 CV is 0.97 mL. All fits obtained an R^2 of 0.999 and an RMSE of 0.08, 0.11, 0.11, and 0.09 for chymtrypsinogen, cytochrome C, lysozyme, and RNase respectively.

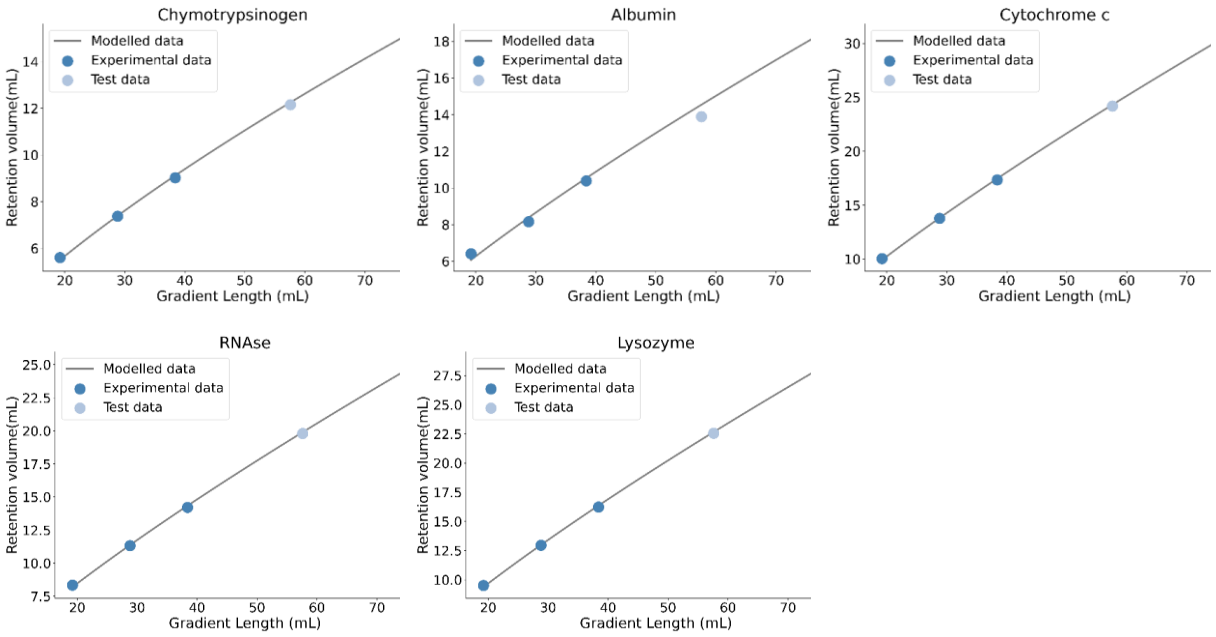


Figure 5.B.2. Fitted regression curves at pH 4.3 (grey line) of the experimental data (dark blue dots) and the test data point (light blue dot) at 58.2 mL, equal to 60 CV as 1 CV is 0.97 mL. All fits obtained an R^2

S

of 0.999 and an RMSE of 0.07, 0.22, 0.10, 0.10, and 0.09 for albumin, chymtrypsinogen, cytochrome C, lysozyme, and RNase respectively.

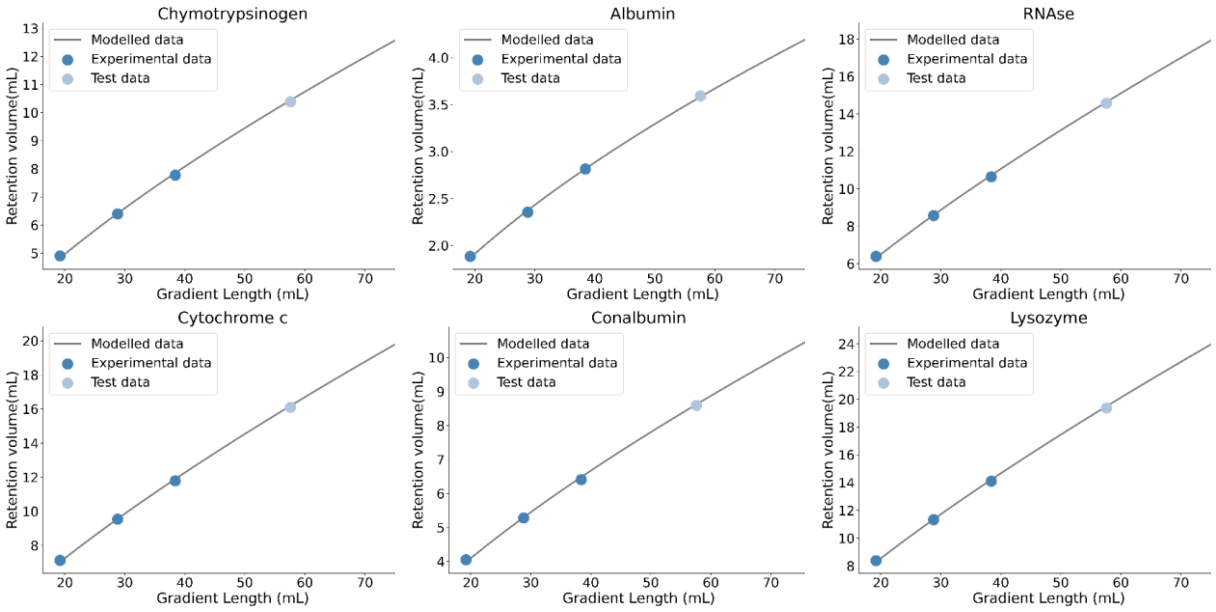


Figure 5.B.3. Fitted regression curves at pH = 5.0 (grey line) of the experimental data (dark blue dots) and the test data point (light blue dot) at 58.2 mL, equal to 60 CV as 1 CV is 0.97 mL. All fits obtained an R^2 of 0.999 and an RMSE of 0.01, 0.05, 0.06, 0.06, 0.07, and 0.08 for albumin, chymotrypsinogen, cytochrome C, lysozyme, RNase, and conalbumin respectively.

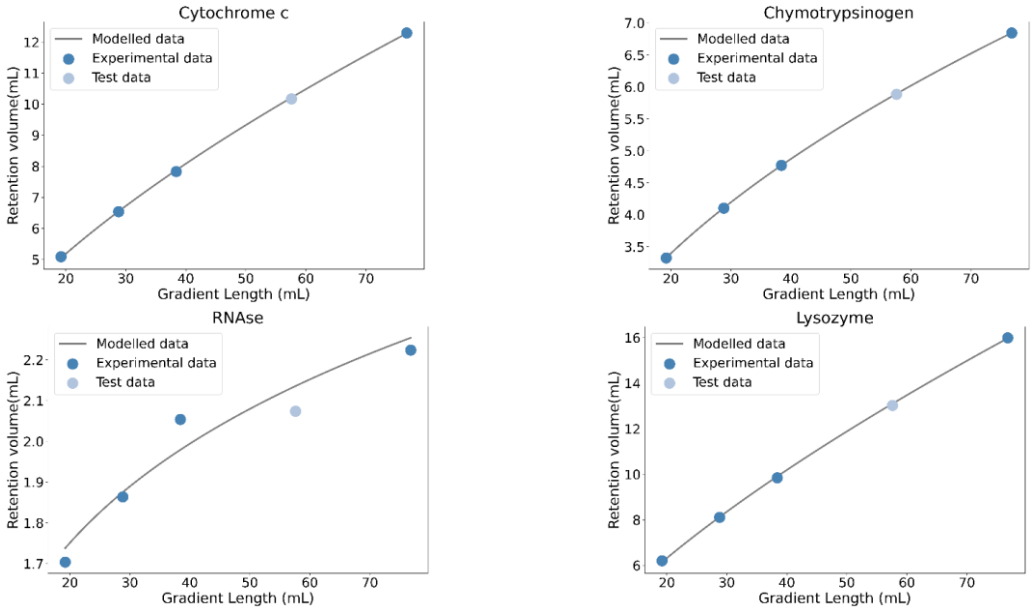


Figure 5.B.4.. Fitted regression curves at pH 7.0 (grey line) of the experimental data (dark blue dots) and the test data point (light blue dot) at 58.2 mL, equal to 60 CV as 1 CV is 0.97 mL. All fits obtained an R^2 of 0.999, except for RNAse that has an R^2 of 0.95. The RMSE values are 0.03, 0.002, 0.04, and 0.04 for cytochrome C, chymtrypsinogen, RNAse, and lysozyme respectively.

Appendix 5.C

Additional data for the mechanistic model validated at pH 7.0. For all proteins at pH 7.0, the maximum retention peak difference is 1.01 CV and the average difference is 0.86 CV, which is 1.68% and 1.43% with respect to the gradient length (60 CV). To assess the concentration agreement between the modeled and experimental results, we compared the difference between the peak width at half of the peak maximum and the peak concentration. RNAse was left out of this comparison for the peak width difference, as determining half of the peak maximum is not possible for the experimental data. The maximum peak width difference is 2.07 CV, equal to 2.23% relative to the gradient length (60 CV). The average peak width difference is 0.81 CV, equal to 1.35% relative to the gradient length (60 CV). The peak concentration differs maximally by 0.04 mg/mL, which deviates about 7.8% to the initial concentration. The average difference in the peak concentration is 0.01 mg/mL, equal to 3.1% relative to the initial concentration.

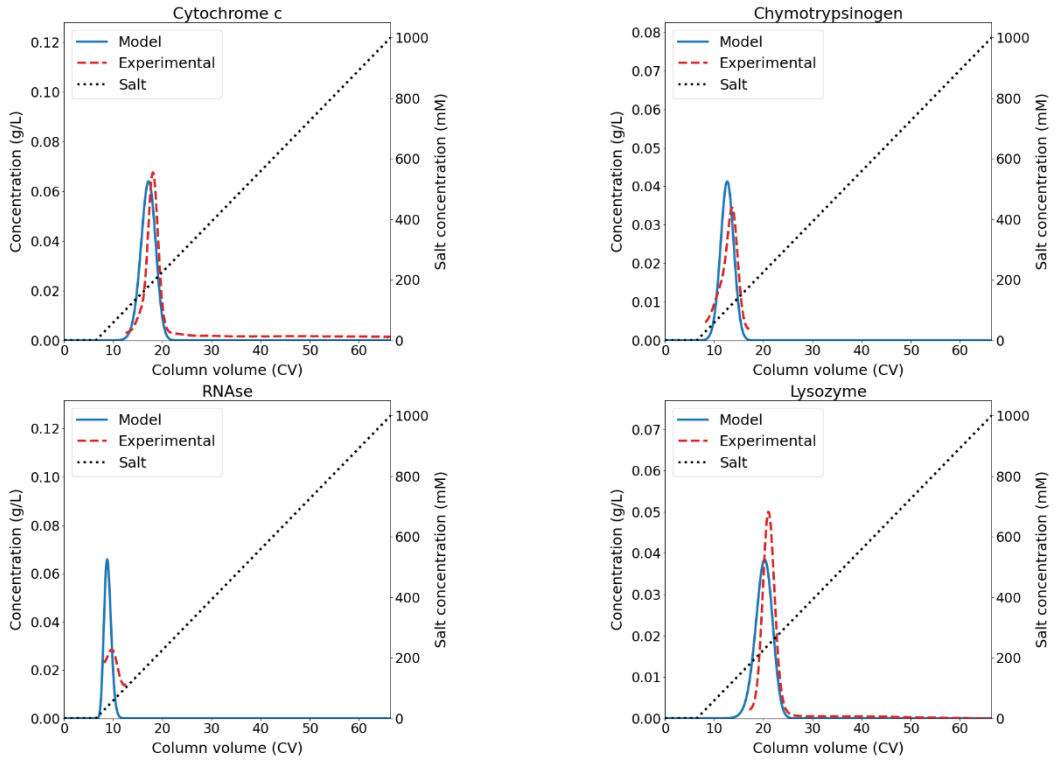


Figure 5.C.1. Chromatographic mechanistic model validation for gradient length of 60 CV, equal to 58.2 mL, at a pH of 7.0. Blue line indicates the MM predicted concentration of the protein, while the red dotted line indicates the experimental concentration. The black dotted line indicates the salt concentration. The initial concentrations are chymotrypsinogen: 0.46 mg/mL, cytochrome C: 0.80 mg/mL, lysozyme: 0.55 mg/mL, and RNase: 0.39 mg/mL.

Appendix 5.D

Calibration lines for each protein at pH 5.0 and 7.0, shown in Figure 5.D.1 and 5.D.2 respectively.

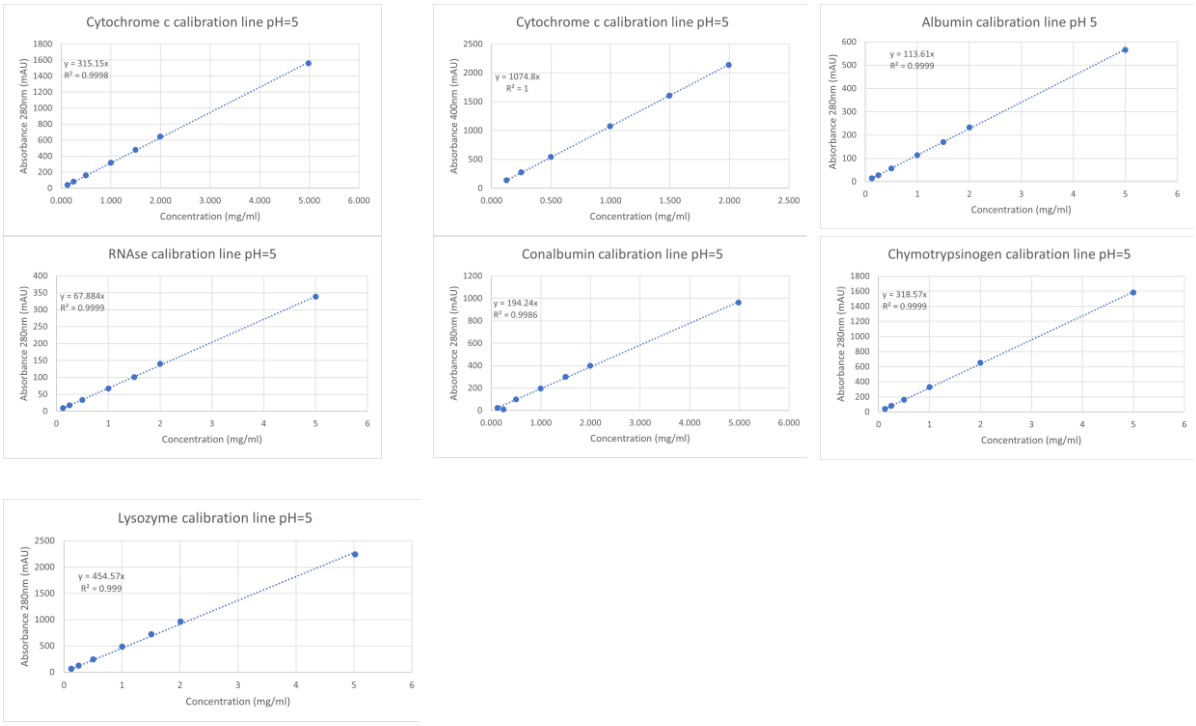


Figure 5.D.1. Calibration lines (blue dotted line) for each protein at pH 5, the blue dots indicate the experimental data. The concentrations are measured at an absorbance of 280 and 400 nm. 400 nm absorbance is specifically needed to quantify cytochrome C.

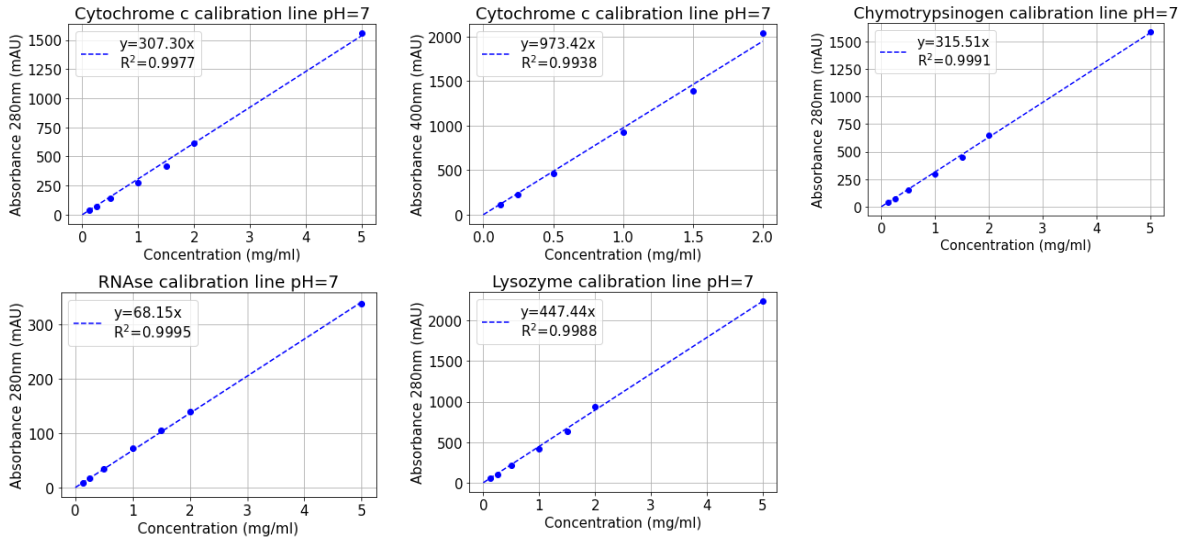


Figure 5.D.2. Calibration lines (blue dotted line) for each protein at pH = 7.0, the blue dots indicate the experimental data. The concentrations are measured at an absorbance of 280 and 400 nm. 400 nm absorbance is specifically needed to quantify cytochrome C.

Appendix 5.E

The consistency of the optimization case study was evaluated by running the same optimization five times. The QSPR-based and experimental-based method results are shown in Table 5.E.1 and 5.E.2 respectively.

Table 5.E.1. Optimization results using the QSPR-based method, showing the performance measurements and obtained optimized variables. K_{eq}= 0.028 and v = 3.05.

	Purity (%)	Yield (%)	HCP clearance (%)	Product concentration (g/L)	Lower cut point (%)	Upper cut point (%)	Initial salt concentration (mM)	Final salt concentration (mM)
1	74.33	97.50	79.79	0.32	4.40	91.70	14.80	330.40
2	73.66	97.81	79.01	0.30	3.70	92.80	19.80	324.50
3	73.91	97.68	79.31	0.30	4.20	93.00	24.40	327.90
4	74.23	97.48	79.69	0.34	4.70	92.30	17.70	354.70
5	74.44	97.40	79.93	0.31	4.40	90.90	18.00	325.90
Maximum difference	0.78	0.41	0.92	0.03	0.90	2.10	9.60	30.20

Table 5.E.2. Optimization results using the experimental-based method, showing the performance measurements and obtained optimized variables. K_{eq}= 0.071 and v = 2.37

	Purity (%)	Yield (%)	HCP clearance (%)	Product concentration (g/L)	Lower cut point (%)	Upper cut point (%)	Initial salt concentration (mM)	Final salt concentration (mM)
1	74.63	96.30	80.36	0.30	7.69	91.21	24.54	320.58
2	74.09	96.62	79.72	0.29	8.54	91.78	22.14	320.00
3	74.22	96.50	79.88	0.29	8.32	91.91	36.47	321.72
4	74.45	96.44	80.15	0.30	8.54	90.80	23.90	320.85
5	74.59	96.38	80.30	0.30	7.99	91.94	28.55	320.13
Maximum difference	0.50	0.23	0.58	0.01	0.85	1.14	14.33	1.72

List of abbreviations

Abbreviation	Definition
AEX	Anion exchange chromatography
AI	Artificial intelligence
ANN	Artificial neural networks
BSA	Bovine Serum Albumin
CEX	Cation exchange chromatography
CPP	Critical process parameters
CQA	Critical quality attributes
CV	Column volume
D	Dilution
DF	Diafiltration
DoE	Design of experiments
EP	Electrostatic potential
HCP	Host cell protein
HIC	Hydrophobic interaction chromatography
HT	High throughput
HTE	High throughput experimentation
HTPD	High throughput process development
IEX	Ion exchange chromatography
LGE	Linear gradient experiments
LHS	Liquid handling station
mAbs	Monoclonal antibodies
MLR	Multi linear regression
MM	Mechanistic modeling
Nan	Not-a-number
ODE	Ordinary differential equation
OFAT	One-factor-at-a-time
PAT	Process analytical technologies
PDE	Partial differential equation
QbD	Quality by design
QSAR	Quantitative structure activity relationship
QSPR	Quantitative structure property relationship
ReLU	Rectified linear unit
RL	Reinforcement learning
RMSE	Root mean squared error
RSM	Response surface methodology
TFF	Tangential flow filtration
UF	Ultrafiltration
UF/DF	Ultrafiltration and diafiltration

UFVVD	ultrafiltration with variable volume diafiltration
VLP	Virus-like particles
WOP	Weighted overall performance

Curriculum vitae

Daphne Keulen was born on May 9th, 1994, in Geldrop, The Netherlands. She spent her childhood in the southern part of the Netherlands, in the small village of Heeze. After completing high school, she moved to Leiden for her studies, and is currently living in The Hague.

During high school, she already showed a strong interest in beta-courses, particularly mathematics and biology. Consequently, pursuing a degree in Life Science and Technology, offered jointly by Leiden University and the Technical University of Delft, seemed like a natural choice. As she progressed through her bachelor's degree, she discovered a stronger affinity for engineering courses. This strengthened her decision to pursue her master's degree in Life Science and Technology at TU Delft, where she specialized in Bioprocess Engineering, a field she thoroughly enjoys. During her Erasmus exchange in Copenhagen at the Technical University of Denmark, she broadened her understanding of various subjects, including life cycle assessment of products and systems, as well as strategy and planning methods. Throughout her master's thesis, supervised by Marcel Ottens, she developed a great interest in mathematical modeling and downstream process engineering, particularly for chromatography. After completing her master thesis, she did an internship at Evides Industrial Water, where she developed a strategic plan outlining and prioritizing their research projects for the upcoming years.

In 2019, she started her PhD at the Bioprocess Engineering department with Marcel Ottens and Martin Pabst as her promotor and co-pomotor, respectively. This PhD project was part of a larger collaboration with GlaxoSmithKline, Rixensart, supervised by Geoffroy Geldhof and Olivier Le Bussy. Her research focused on developing mathematical mechanistic models and artificial neural networks for downstream process unit operations. These models were used to *in-silico* optimize biopharmaceutical downstream processes. During her PhD, she was a board member of YoungNBV, the Dutch biotechnology association for young professionals. In this role, she was highly involved in organizing networking events and handling organizational tasks for the association.

Acknowledgements

They say: Time flies when you're having fun. Despite facing various challenges along the way, I will always look back on this period as a wonderful time of personal and academic growth. None of it would have been possible without the endless support of my colleagues, friends, and family.

First of all, I want to thank **Marcel Ottens** for his supervision and continuous support throughout these past 4.5 years and before during my master thesis, where I discovered my interest for downstream processing. It has been a great pleasure working with you and I learned a lot from you; from addressing complex project challenges to handling conflicts diplomatically. Your positive attitude and confidence in my abilities have fostered my growth and enable me to thrive.

I want to thank **Martin Pabst,** my co-promotor for his supervision of my project and the valuable feedback he provided. I truly appreciated our meetings, your positive and open-minded attitude, and your critical thinking on my topic that was beyond your research area.

This PhD project would not have been possible without the support of GSK, especially thanks to my supervisors **Geoffroy Geldhof** and **Olivier le Bussy**. Thank you for your industrial insights and guidance throughout the project, your critical industrial perspective often provided fresh insights, complementing our academic approach. I will miss our biweekly meetings and our visits to the GSK site in Belgium. I would also like to thank my other colleagues at GSK for their support in advancing our project.

Roxana, my first buddy who joined the GSK project. I greatly want to thank you for these 4.5 years, without you it would have been much less fun. We learned a lot from working with each other and working with others. Chit-chatting and endlessly talking about the project in our office was sometimes more appealing than doing the actual work. I will remember with great pleasure all the enjoyable and enlightening trips we made together, from our six-week stay in Brussels to our visit to San Francisco for the ACS conference. I greatly enjoyed working with you as a and all the fun we had!

Tim, you made our GSK team complete. You brought the chit-chatting in our office to a higher level, and by that I mean lasting longer. I love the enthusiasm you tell your stories or explaining topics. We had a lot of fun together in the office, before we knew it, it was often already time for coffee break. In the beginning it felt like I didn't know anything about programming and computers, you were so knowledgeable, through the years I learned a lot from you. It really was a pleasure working with you, you're patient and have a positive vibe, thank you a lot for these years working together and becoming a great friend.

During my project, several master students performed their master end project with me. **Myrto**, **Erik**, **Anne-Marijn**, **Manto**, and **Eleni**, thank you all the great time working together. Each of you faced your own challenges and were able to overcome those, it was a pleasure to see you growing throughout the project and being part of that. Thank you all!

Furthermore I want to thank my dear friends and colleagues **Tiago** and **Mariana** for the great fun in the Ottens (and Otters ACS) group. Since the start of my PhD, you both have been my companions, enjoying memorable trips, productive meetings, competitions, discussions about our projects, and not to forget the drinks and dinners we had. **Tiago**, I really enjoyed teaching TLS with you and working together on the mechanistic modelling. **Mariana**, my co-secretary, we can all learn from the queen's attitude, yet it's our gossip sessions and chats that I cherish the most. .

Tim Nijssen, having you as my office mate for about two years was truly a pleasure, your happiness made my days. I am grateful for the enlightening discussion and the inspiration you provided for my fourth chapter. Your patience, calmness, and wise advices were very valuable. Our time together was not only productive but also filled with enjoyable moments, both within the office and outside, and our memorable trip to Porto. Thanks to my other great and fun office mate, **Ramon**. I enjoyed our shared passion for playing the piano and immersing ourselves with the compositions of Chopin and other great composers. It's a pity that I didn't have the time in my final year to play more piano duets together. You positivity, creativity, and critical thinking are admirable, thank you for this great time.

A special thanks to **Song**, with whom I have been working together with since my master thesis. You are truly 'one-of-a-kind', the resilience, always helping others, and be able to make things work in the lab. Your support and assistance to my master students throughout their projects are invaluable. Beyond your technical expertise, you have become a true friend of mine, someone who you trust and rely on. Thank you immensely!

I want to thank all the BPE members. Starting with the old crew; **Monica**, **Debby**, **Bianca**, **Joana**, **Chemna**, and **Rita**, it is because of you that I started my PhD in the BPE group, thank you all! **Marina** for all the great chats we had and our holidays to Mexico for Monica's wedding, it was always fun with you. To **Lars**, who joined this PhD journey with me those past years, together we struggled and enjoyed the beauty of doing a modelling project. **Marijn** and **Oriol**, the unforgettable MES team. I admire your enthusiasm and ability to completely immerse yourselves in a topic. I hope you both regain this characteristic with a positive attitude. I will look back at a great time in San Francisco with you **Maarten**, we laughed a lot, thanks for all the fun and being a great colleague. I want to thank the rest of the great/wonderful/funny BPE group: **Eduardo**, **Hector**, **Joan**, **Meryl**, **Mariana**, **Tamara**, **Brenda**,

Marika, **Mona**, **Miki**, and **Rik**. A big thanks to the BPE staff for their continuous support during my master and PhD time at BPE: **Stef**, **Kawieta**, **Christiaan**, and **Adrie**. Special thanks to **Max**, sharing the office with you during your last year of employment was very special and a lot of fun. **Luuk**, thank you for organizing the 'BPE Ireland studytrip' with Marina and me, it is unfortunate that due to COVID19 we could not go. I hope there will be an opportunity for the BPE group in the future. To **Ludo**, for the great talks and fun. **Marieke**, I am happy you joined TLS teaching with Tiago and me, it was great to get involved with the course on a deeper level and helping you to reorganize the course. Moreover, I want to thank you for your support on the fifth chapter of my thesis. Besides, we enjoyed San Francisco together, NBV events, and many more. To **Cees**, who supported me and my students with modelling questions, I liked our talks and the organization of the Captains dinner.

My dear brothers, **Bart** and **Luuc**, I want thank you for being my paranimphs on this important and special day. Throughout these past years, I could always rely on you both whenever I needed to discuss my struggles during my PhD. You always listened, offered valuable advice, and restored my confidence and courage. Bart, thanks to your guidance, my interest in the field of machine learning arose. You taught me the fundamentals and assisted me with the initial design of my mechanistic model using object-oriented coding. Luuc, as my older brother, your experience navigating PhD challenges and with your pragmatic perspective, you helped me to see the bigger picture and renew my energy to keep going.

To all my friends, I want to thank you for your endless support during these challenging and sometimes seemingly hopeless times. You always listened to my 'very difficult and though' PhD life, it was a pleasure to complain with you. But most importantly, thank you for the moments of joy and laughter we shared together. Thanks to my hockey-friends, Jan 18, Moustache, and my friends from lower and high school. **Vera**, a special thanks to you, over these past months we have worked together on the cover for my thesis, and I love it! I want to thank you for this, for the creative process, and above all as a great friend.

'Mijn rots in de branding' and dear love **Maarten**, thank you for your endless support, encouragement, and to always unburden my heart. We continue to grow and become stronger as a team and we will continue to do so. I am really looking forward to our next adventure, Copenhagen.

'Lieve **mama**', I can't thank you with words, only with a big hug! The past few years have not been the easiest, but we have also been able to experience many special moments and enjoy life again. You understand and know me like no other, and I know that you will always support me. Regarding the PhD period, I was lucky that I could work 'lekker bij mama' especially during COVID and during my writing period, thanks for the delicious food you cooked.

Moreover, I want to thank my sisters-in-law, **Esther** and **Sahar**, and of course my cute little niece, **Sarah**, 'het zonnetje in huis'. To my beloved family in-law, **Jacqueline**, **Reinier**, **Robert** and **Zita**, thank you very much for your support and warmth.

In Memoriam:

I would like to thank my dear dad for his support from above. Thanks to the endless trust you have always given me since my childhood, I was able to face and overcome the toughest challenges during my doctoral research. Because of you, I learned the basic knowledge of what I know today, from chemistry in high school to thermodynamics during my bachelor's. I could always rely on your help; you had so much knowledge, I admire this and it inspires me to keep on learning. Thanks to your support, I maintained my perseverance and was able to complete my doctorate.

I know you would be the proudest of all.

List of publications

Journal publications

D. Keulen, E. van der Hagen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Using artificial neural networks to accelerate flowsheet optimization for downstream process development, Biotechnology and Bioengineering (2023) 1-14. https://doi.org/https://doi.org/10.1002/bit.28454.

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

M. Moreno-González, D. Keulen, J. Gomis-Fons, G.L. Gomez, B. Nilsson, M. Ottens, Continuous adsorption in food industry: The recovery of sinapic acid from rapeseed meal extract, Sep Purif Technol 254 (2021) 117403. https://doi.org/https://doi.org/10.1016/j.seppur.2020.117403.

M. Moreno-González, V. Girish, D. Keulen, H. Wijngaard, X. Lauteslager, G. Ferreira, M. Ottens, Recovery of sinapic acid from canola/rapeseed meal extracts by adsorption, Food Bioprod Process 120 (2020) 69-79. https://doi.org/https://doi.org/10.1016/j.fbp.2019.12.002.

Conference contributions

D. Keulen, E. van der Hagen, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Flowsheet optimization for biopharmaceutical processes. 14th European Congress of Chemical Engineering and 7th European Congress of Applied Biotechnology (ECCE-ECAB), Berlin, Germany, September 2023, oral presentation

D. Keulen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Comparing in silico optimization strategies to develop biopharmaceutical downstream processes. American Chemical Society (ACS) Fall meeting, San Francisco, United States, August 2023, oral presentation.

D. Keulen, E. van der Hagen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Using artificial neural networks to accelerate flowsheet optimization for downstream process development. GSK Vaccines R&D Days, Rixensart, Belgium, June 2023, poster presentation.
D. Keulen, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Optimizing biopharmaceutical downstream processes using neural networks and mechanistic models. Biopartitioning and purification conference (BPP), Lisbon, Portugal, September 2022, oral presentation.

D. Keulen, M. Apostolidi, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Towards efficient chromatographic modeling of complex vaccine mixtures. Vaccine Technology VIII, Sitges, Spain, June 2022, poster presentation.

D. Keulen, M. Apostolidi, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Towards efficient chromatographic modeling of complex vaccine mixtures. Netherlands Process technology Symposium (NPS), Delft, The Netherlands, April 2022, poster presentation.

D. Keulen, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Model-based process development for subunit vaccines. 16th International PhD Seminar on Chromatographic Separation Science (SoCSS), Vienna, Austria, July 2021, oral presentation.

D. Keulen, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Towards efficient chromatographic modeling of complex vaccine mixtures. 13th European Congress of Chemical Engineering and 6th European Congress of Applied Biotechnology (ECCE-ECAB), online, September 2021, poster presentation.

D. Keulen, R. Disela, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens*, In-silico* vaccine purification process development. 13th European Symposium on Biochemical Engineering Sciences (ESBES), online, May 2021, oral presentation.

D. Keulen, R. Disela, M. Pabst, M. Ottens, Rational and systematic purification process development – the next generation. The 5th international high-throughput process development (HTPD) meeting, Porto, Lisbon, November 2019, poster presentation.

D. Keulen, G. Geldhof, O. Le Bussy, M. Pabst, M. Ottens, Rational and systematic vaccine purification process development. GSK Vaccines R&D Days, Rixensart, Belgium, November 2023, Flash talk.