Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Doctoral Dissertation

Semi-automation of bioanalyses for real-time monitoring of quantity, purity and activity of biopharmaceuticals

submitted by

DI[in] Anna CHRISTLER, BSc

in partial fulfilment of the requirements for the academic degree

Doktorin der Bodenkultur (Dr.nat.techn.)

Vienna, November 2021

Supervisor:
Univ.Prof. DI Dr. nat. techn. Alois Jungbauer
Institute of Bioprocess Science and Engineering
Department of Biotechnology

Affidavit (Eidesstattliche Erklärung)

I hereby declare that I have authored this dissertation independently, and that I have not used any assistance other than that which is permitted. The work contained herein is my own except where explicitly stated otherwise. All ideas taken in wording or in basic content from unpublished sources or from published literature are duly identified and cited, and the precise references included. Any contribution from colleagues is explicitly stated in the authorship statement in the published papers.
I further declare that this dissertation has not been submitted, in whole or in part, in the same or a similar form, to any other educational institution as part of the requirements for an academic degree.
I hereby confirm that I am familiar with the standards of Scientific Integrity and with the guidelines of Good Scientific Practice, and that this work fully complies with these standards and guidelines.

Vienna, November 2021 Anna CHRISTLER *(manu propria)*

If I have 1000 ideas and one of them turns out to be good, I am satisfied.

Alfred Nobel

Acknowledgements

I want to thank my supervisor Alois who is one of the most open-minded, lively, and competent people I have met. He encouraged me to pursue my doctoral degree and has been true to his word to push me forward whenever I was stuck on the way.

I also thank my advisor Astrid Dürauer for her support during my thesis.
I thank Dominik Sauer, Nicole Walch, Edit Felföldi, Magdalena Mosor and Anna Frank for discussing (strange) results with me and for the fun we had in the lab and on conferences! I thank Didi Palmberger for his mental and active support during my time at BOKU and acib.

I acknowledge the exceptionally good collaboration with the project team members from our industrial partners Böhringer Ingelheim Regional Center Vienna and Sandoz/Novartis. The quarterly discussions aligned our efforts to industrial interests as well as fundamental research on protein binding and analysis.

I am grateful for the support from acib in hosting the 16th International PhD Seminar on Chromatographic Separation Science (SoCSS). It was quite a challenge to organize the event and Verena Beck, Patricia Velikogne, Astrid Preisz and Matthias Drexler made indispensable contributions to its success.

Finally, I thank my parents and my sister for their continued support and confidence.

The research was conducted with the support of the Austrian Federal Ministry for Digital and Economic Affairs (bmwd), the Federal Ministry for Transport, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, the Government of Lower Austria and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. The funding agencies had no influence on the conduct of this research. Additional financial support was provided by the industrial partners.

Abstract

Real-time monitoring is the basis for real-time batch release concepts and automation in biopharmaceutical manufacturing. Challenges of real-time monitoring systems with model-based prediction of quality attributes are the experimental expenditure for model training and the transferability of the models to different sites. A real-time monitoring system based on process analytical technology (PAT) and data-based models was developed by Walch et al. 2019 and Sauer et al. 2019 to estimate quality attributes of biopharmaceuticals during chromatography. In this thesis, data generation for model training was semi-automated using a liquid handling station. I hypothesized that semi-automated analytics is more precise and less time consuming than manual analysis. The time to analyze all fractions of a chromatography run was reduced by 2.4 h. Operator influence was reduced as shown by a 48% decrease in variability in DNA quantification. Precision of the semi-automated methods was equal to manual methods based on the comparison of results generated by one operator. Furthermore, I hypothesized that the prediction models are transferable to different sites. Root mean squared errors of predictions at the new sites were on average twice as high as at the training site. This was attributable to limited sensor robustness of the fluorescence detector which was a prototype and to model extrapolation. At all sites, prediction of purity was more challenging compared to product concentration. Model training for new processes at the new sites allowed accurate real-time monitoring with the transferred systems. The use of automated methods unburdens operators from repetitive and potentially hazardous tasks. Process monitoring by PAT is inherently faster than by offline analysis and more specific than conventional monitoring of elution by UV absorption. Compared to conventional “offline” manual workflows, different skills are required when using automated systems.

Deutsche Zusammenfassung

Echtzeit-Überwachung von Herstellprozessen ist die Grundlage für Echtzeit-Chargenfreigabe und Automatisierung in der biopharmazeutischen Industrie. Herausforderungen solcher Systeme mit modellbasierter Vorhersage von Qualitätsmerkmalen sind der experimentelle Aufwand für die Modellentwicklung sowie die Übertragbarkeit der Modelle an andere Standorte. Ein Echtzeit-Überwachungssystem basierend auf Prozessanalysetechnologie (PAT) und datenbasierten Modellen wurde von Walch et al. 2019 und Sauer et al. 2019 für chromatographische Trennprozesse entwickelt. In der vorliegenden Doktorarbeit wurde die Datenerzeugung für das Modelltraining mit einem Pipettierroboter teil-automatisiert. Ich prüfte die Hypothese, dass teil-automatische Analysen präziser und weniger zeitaufwendig als manuelle Analysen sind. Die erforderliche Zeit der Analytiker wurde um 41% reduziert, die Präzision der teil-automatischen Methoden verbesserte sich allerdings nicht. Weiters stellte ich die Hypothese auf, dass die Vorhersagemodelle auf neue Standorte übertragbar sind. Der mittlere Vorhersagefehler an den neuen Standorten war im Durchschnitt doppelt so hoch wie am ursprünglichen Standort. Dies war auf die eingeschränkte Robustheit des Fluoreszenzdetektors, ein Prototyp, und auf teilweise Modellextrapolation zurückzuführen. An allen Standorten war die Vorhersage der Reinheit mit größeren Fehlern verbunden als jene der Produktkonzentration. Modellentwicklung für neue Prozesse an den neuen Standorten ermöglichten eine genaue Echtzeitüberwachung mit den transferierten Systemen. Der Einsatz automatisierter Verfahren entlastet die Bediener von sich wiederholenden und potenziell gefährlichen Aufgaben. Die Prozessüberwachung durch PAT ist von Natur aus schneller als die Offline-Analyse und spezifischer als die konventionelle Überwachung der Elution durch UV-Absorption. Beim Einsatz automatisierter Systeme sind im Vergleich zu herkömmlichen, manuellen Methoden andere Fähigkeiten erforderlich.

Abbreviations

Abbreviation	Explanation
API	active pharmaceutical ingredient
BSA	Bovine serum albumin
CHO	Chinese Hamster Ovary (cells)
cGMP	current Good Manufacturing Practices
CPP	Critical process parameter
CQA	Critical quality attribute
DF	Dilution factor
DoE	Design of experiments
dsDNA	Double-stranded DNA
E. coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
FDA	Food and Drug Administration
FGF-2	Fibroblast growth factor 2
HMWI	High molecular weight impurities
HPLC	High pressure liquid chromatography
HCP	Host cell proteins
i.e.	*lat. id est,* that is
IEX	Ion exchange chromatography
IgG	Immunoglobulin G
IR	Infrared (spectroscopy)
kDa	kiloDalton
LLOQ	Lower limit of quantification
LMWI	Low molecular weight impurities (fragments)
mAb	Monoclonal antibody
MALS	Multi-angle light scattering
MIR	Mid infrared
NIR	Near infrared
PAT	Process Analytical Technology
QbD	Quality by Design
QTP	Quality Target Profile
RFID	Radio-frequency Identification
RI	Refractive index
RMSE	Root mean squared error
RSD	Relative standard deviation
ULOQ	Upper limit of quantification
UPLC	Ultra-high pressure liquid chromatography
UV/VIS	Ultra-violet/visible light (absorption spectroscopy)
WHO	World Health Organization

Table of Contents

1. Introduction

1.1. Real-time monitoring in biopharmaceutical manufacturing

Economic competition and product diversification of biopharmaceuticals requires their production processes to be developed faster, cheaper and under highest quality requirements [1], [2]. For almost 20 years, the regulatory agencies have encouraged the pharmaceutical companies to replace the conventional empirical approach for product and process development by a rational approach. They named it Quality by Design (QbD): ‘‘a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management’’ [3]. However, the pharmaceutical industry with its stringent regulatory requirements is conservative and implementation is still ongoing. The slow progress is mainly due to the high complexity of biotechnological products and manufacturing processes which are still not fully understood. Many decisions are still made based on experience. Machine learning and artificial intelligence are needed to achieve the state of control already in use in other industries [4].

One enabling technology on the way to QbD is real-time process monitoring where critical quality attributes such as product concentration and purity are monitored in real-time or near real-time [5]. In conventional processes, product quality attributes are monitored by offline analyses in the analytical laboratory. A hold time is required when decisions about how to continue the process are made based on the lab results. Real-time monitoring allows for faster decision-making and a reduction of the total processing time. Furthermore, continuous production, which has the potential of increased productivity with reduced footprint, is only possible with the use of monitoring systems [6].

Real-time monitoring systems are also called process analytical technology (PAT) [7] or soft sensors (from *software*) [8]. The use of this technology is encouraged by regulatory authorities [5] because a more active control of product quality is enabled compared to conventional “retrospective” quality control as described before. Moreover, easier and faster process optimization is possible because changes within the filed design spaces are not considered as process changes [9]. Also, quality risk analysis can be based on process understanding and/or representative process models. PAT solutions have been developed and are in use in other industries for decades. Continuous monitoring of manufacturing steps, so-called *unit operations*, is the basis for process control during manufacture. Controlled processes enable manufacturers to produce (more) constant output from variable input. This is especially relevant for biotechnological products where the biological system is an inherent source of variability in the manufacturing process.

From an economic but also from the quality point of view of a manufacturer, a vision is real-time batch release [10], [11] (Figure 1). Real-time release means that the decision about release or rejection of a batch is solely based on data from online measurements. The benefits are substantial reductions in time, cost, and risk. The conventional batch release procedure based on finished product testing has a small but inherent statistical risk of approving a batch although it doesn’t fulfill specifications. In the words of one of the world’s most rigorous regulatory agency, the American Food and Drug Administration (FDA): “(…) quality cannot be tested into products; it should be built-in or should be by design” [5]. The European Medicines Agency (EMA) states in their Guideline on Real Time Release Testing [10] that “under specific circumstances an appropriate combination of process controls (critical process parameters) together with pre-defined material attributes may provide greater assurance of product quality

than end-product testing". Additional to enabling proactive quality control, real-time release is faster compared to finished product testing.

Real-time batch release must be approved by the concerned authorities for each product and production site. Approval is based on the demonstration of process control which is based on product knowledge, process understanding and a solid quality (risk) management system (see Figure 1). Recently, so-called "digital twins" of the production processes or even facilities are being named as an enabling technology to meet current good manufacturing practice guidelines (cGMP) and economic requirements [12].

Figure 1: Fundamental requirements for real-time batch release.

1.2. *Chromatography in downstream processing and critical product quality attributes*

Chromatography is contained as a unit operation in almost every downstream process of biotherapeutics due to its unrivaled resolution. The separation of the components in the feed material is based on their interaction with a stationary phase. In most cases, the stationary phase is equipped with ligands to which the sample components bind differentially. The interaction can be based on, for example, affinity binding, ion exchange, hydrophobic interaction or, in size exclusion chromatography, on the size of the components [13]. Thus, the sample components are retarded to different degrees by the stationary phase and ideally also eluted at different times to achieve the desired separation. Elution is most often caused by a buffer of different ionic strength or pH or both. The product must then be collected in a certain time interval to ensure the desired separation from the impurities. The collection step is called *product pooling*. The eluent stream is often collected in fractions which are then analyzed, and those fractions pooled which result in desired quality attributes. Product pooling is performed in the industry by either one of two methods [14]: When relatively small amounts of impurities are present or resolution of product and impurities is good, a single fraction is collected. The start and end points of collection are usually determined based on inline measurement of UV absorption, often at 280 nm. In contrast, when relatively large amounts of impurities are present or resolution is limited or online measurements cannot capture the quality attributes, the column effluent is collected in equally sized fractions, and the quality target profile of all fractions is analyzed in a wet lab before pooling the fractions which give the desired quality attributes (e.g., purity, quantity, yield).

Typical quality attributes defined in the product quality target profile (QTP) are product concentration (also referred to as *quantity*), purity and activity. Impurities are divided into

process-related and *product*-related impurities. Common *product*-related impurities of therapeutic proteins are aggregates, fragments, and variants such as charge variants. *Process*-related impurities derive from the expression system (the “host cells”), the feed medium and downstream unit operations. Common process-related impurities are host cell DNA, host cell proteins (HCP), endotoxins (for products from *E.coli* and other gram-negative bacteria), leached protein A (for monoclonal antibodies), residual solvents or additives and viruses (for products from mammalian expression systems). All impurities must be reduced below levels which are safe for the patient. These levels are specified by regulatory health authorities in their respective pharmacopoeias. *Activity* or *efficacy* describes the ability of the product to perform its intended action which is most often the binding to its specific target (ligand), for example an antigen or a receptor. Therefore, correct folding is necessary for proteins. In many cases, posttranslational modifications (PTMs) are also required for (full) protein functionality. All critical quality attributes (CQA) must be defined and measured with appropriate methods to ensure that the specifications are being met.

The Quality by Design concept also requires the identification of critical process parameters (CPP) which directly influence the product’s CQAs. The CPPs must be monitored and controlled. CPPs in chromatography can be for example the conductivities and pH values of buffers and load materials, the concentrations of product and impurities in the load, the flowrates in the process steps or the column loading [15]. Upper and lower boundaries need to be defined for each CPP which results in multi-factorial design and control spaces.

Determining the pooling start and end time points after preparative chromatography is a multi-objective or “Pareto” optimization problem because the three objectives “yield”, “purity” and “productivity” are in conflict with each other (e.g. [16]). Knowledge about further downstream operations is used to define the thresholds for each attribute. For example, impurities which can be efficiently reduced downstream, can be relatively high in the previous steps. In this light, the power of a model system which covers the whole production process becomes obvious.

The first method which was established for detailed online monitoring (*online* here meaning “on the shop floor”, i.e. in the production area) was *at-line* HPLC (e.g. [17]): Samples are drawn automatically and analyzed by HPLC. HPLC measurements provide information of high accuracy and precision, and the method is well established. However, due to common measurement times of 10 – 30 minutes, at-line HPLC can accelerate offline analysis but cannot be used for real-time monitoring. The only real-time monitoring method established in biopharmaceutical industry is based on UV absorbance, as described above, which is unspecific to the compounds in the sample stream. Therefore, fast component-specific real-time monitoring methods are needed.

As an important unit operation in biopharmaceutical manufacturing, lab-scale chromatographic separation processes were used to develop a real-time monitoring system (Figure 2). A battery of online sensors was used, and statistical models were trained to estimate CQAs in real-time to enable automated collection of the product stream.

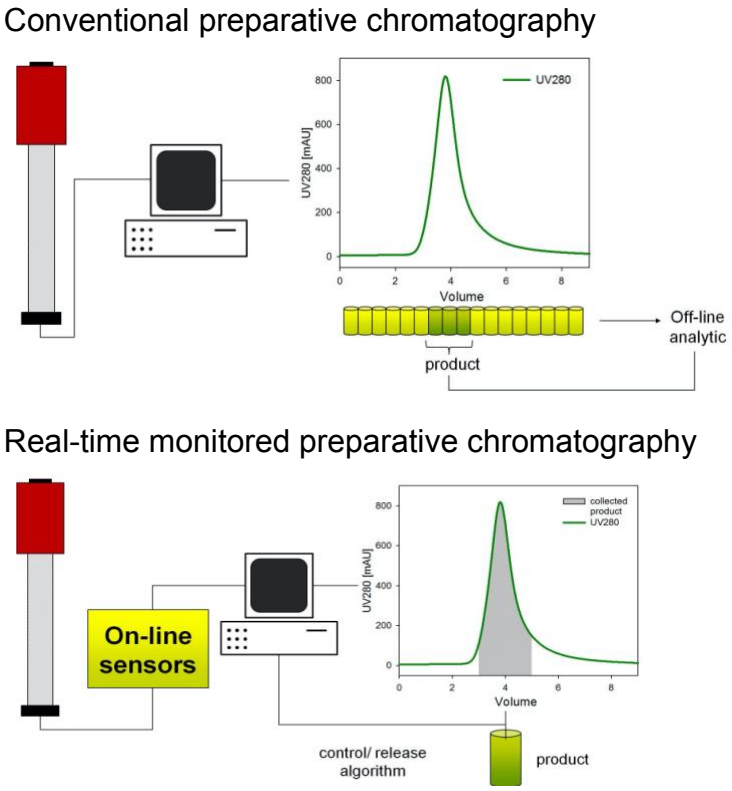


Figure 2: Conventional preparative chromatography using product fractionation and offline analysis for quality determination and real-time monitored chromatography using online sensors and statistical models for automated product pooling. The latter represents the system transferred to two new sites during this thesis. Author: Nicole Walch, printed with kind permission.

1.3. *Model proteins and model processes*

Two industry-relevant recombinant model proteins and processes were used to develop the monitoring system. The recombinant proteins were expressed using two very well-characterized and commonly employed expression systems for therapeutic proteins: *E.coli,* a gram-negative bacterium, and Chinese hamster ovary cells (CHO), a mammalian cell factory system. They are two of the oldest "biotechnological workhorses" and represent two very different classes of expression systems and products.

As a model protein from bacterial fermentation, recombinant basic human fibroblast growth factor 2 (FGF-2) was over-expressed in *E.coli* as described in [18]. FGF-2 is used clinically for example to support tissue repair, in regenerative medicine or for stem cell expansion. It has a molecular weight of around 17 kDa and an isoelectric point (pI) of 9.6. After harvest, homogenization and clarification, a two-step chromatographic purification sequence was developed [18] and used to generate data for training and testing of statistical models for real-time monitoring [19]. For the first chromatographic step in the purification scheme, the so-called "capture" step, anion exchange chromatography and heparin affinity chromatography were compared [18]. The critical quality attributes were: FGF-2 quantity (i.e. concentration), purity – in terms of monomer content (%), high molecular weight impurities (HMWI, %), low molecular weight impurities (LMWI, %), concentrations of host cell dsDNA (ng/ml or ppm), host cell proteins (ng/ml or ppm) and endotoxins (EU/ml) – and the binding affinity to heparin (dissociation constant K_D in nmol/l). High and low molecular weight impurities are product-related impurities which arise from aggregation and fragmentation of the product.

As a model protein from mammalian cell culture (CHO) the monoclonal antibody (mAb) *Adalimumab* was used. Adalimumab is an immunosuppressive IgG1-type antibody inhibiting the immunostimulating cytokine *tumor necrosis factor α* (TNF-α) [20]. Adalimumab is used for the treatment of persistent forms of inflammatory diseases and autoimmune diseases such as different forms of arthritis, psoriasis, gastrointestinal inflammatory diseases. Adalimumab was first supplied to the market as Humira®. Between 2015 and 2019, Humira was the drug with the highest revenues by far[1]. With about 10 biosimilars approved in EU, US and Indian markets together[2], it is an example of the aforementioned increasing market pressure [21]. Adalimumab has a molecular weight of about 144.2 kDa and various posttranslational modifications (e.g. O-/N-glycosylations and disulfide bridges) and can undergo modifications during processing (e.g. methionine oxidation, deamidation, isomerization) leading to a huge microheterogeneity [22]. Adalimumab was purified by Protein A affinity chromatography [23]. The CQAs assessed for Adalimumab were the same as for FGF-2 except for endotoxins since mammalian cells do not produce endotoxins. The binding affinity to the ligand TNF-α was measured [23].

1.4. *Offline wet lab analytics*

Protein concentrations and product-related impurities (HMWI and LMWI) were determined using HPLC or UPLC methods [18], [23]. Process-related impurities (HCP, DNA and endotoxins) and binding affinity were quantified by common biochemical analyses [24].

General requirements to analytical methods are that they are accurate (i.e. measuring the correct value) and precise (i.e. repeatedly measure the same or very similar results). Also, they must be specific for the analyte of interest. This means that the method can detect and quantify the analyte despite the presence of other species. Ideally, the method also has a wide dynamic range, meaning that it can quantify a large window of analyte concentrations.

Equipment and methods should be robust against variations in external conditions such as temperature, humidity as well as variations in materials and users. Measuring control samples without analyte (blank sample) and one or more known concentrations are the standard method to check the method during every set of samples.

If dilution factors greater than 10 are required, samples are usually diluted in several steps, for example 1:5 or 1:10. Dilution errors of each step multiply but precision is statistically better in this way compared to one dilution step with a large dilution factor. When many samples must be diluted in the same or similar way, this is often done in multi-well plates with, for example, 96 wells (8 x 12) of about 400 µl each.[3] Multi-channel pipettes are used to dispense reagents or dilute samples simultaneously.

Most plunger-operated pipettes operate based on volume replacement and use disposable polymeric tips. Tips can be available as especially “low retention” or non-sticky. In the standard pipetting technique, also called “forward” pipetting, the exact volume is aspirated and the liquid in the tip is dispensed first at low speed and the remaining micro- or nanoliters (depending on sample volume and viscosity) are pushed out by a blow-out. The second technique is called “reverse pipetting”, where some extra microliters are aspired which then remain in the tip because no blow-out is performed. The reverse technique is applied mainly for liquids that tend to form bubbles to avoid squirting of the sample to other samples in a dilution plate.

[1] https://en.wikipedia.org/wiki/List_of_largest_selling_pharmaceutical_products, accessed September 01, 2021.
[2] https://en.wikipedia.org/wiki/Adalimumab#Biosimilars, accessed October 04, 2021.
[3] Different plates are available with six to 3456 wells for different purposes, e.g., cultivation of cells and high-throughput analysis, respectively.

A critical process for accuracy and precision of an analysis is sample dilution and the corresponding pipetting technique. Pipetting accuracy and precision are influenced, among others, by:

- The type of pipette (volume range)
- Speed of aspiration and dispense
- The angle of the pipette during aspiration and dispense
- Prewetting of the tip before aspiration of the sample
- Depth of immersion into the sample

For more details on good pipetting practice, compare the infographic in Section 9.1.

Analyzing several or many parameters of numerous fractions per chromatography run by wet chemistry was work- and resource-intense. Manual work for data generation can be reduced by automated and/or semi-automated systems.

1.5. Laboratory automation

Automated and semi- or partially automated methods are used in many different disciplines and industries such as (exemplary references are given) medicine [25], chemical synthesis [26], agriculture [27], wildlife research [28], electronics [29] or literature review/analysis [30]. The aims of using automated systems can be: reduction of (repetitive) manual work, reduced sample analysis or "turnaround" time, reduced costs and/or variability among operators, increased quality and/or working safety [31]. Another advantage of (semi-)automated systems is that they often provide audit trails which allow the tracking of procedures. Barcodes and/or RFID technology for samples, plates and reagents can improve data integrity and sample traceability even further. After implementing an automation strategy, human actions can focus on steps which are non- or less repetitive and/or require intellectual evaluation. This makes workflows safer by avoiding tiring repetitive work steps [32] and more interesting at the same time. In many analytical methods, sample processing steps such as dilution, aliquoting, addition of reagents, mixing, incubation, centrifugation etc. represent the most labor-intensive tasks and are also well suited for automation [33]. Also data analysis and process development can be (partially) automated, e.g. [34].

In bioanalytical and bioprocessing labs, different kinds of automated systems are available and commonly found. Examples are given in Figure 3. Automated systems can be classified by different aspects [33], for example the degree of automation and the complexity of the tasks performed. In this sense, lab automation begins with task-targeted equipment [31] such as electric pipettes, heating and stirring plates, shakers or autosamplers. On the other side of the spectrum, totally automated laboratories exist where high samples volumes must be analyzed in the same way. Such total automation systems are mostly found in clinical laboratories in hospitals where they consolidate several specialized labs into one core laboratory [35]. Total automation systems also contain automated systems for sample storage, archival and retrieval [31]. Hawker estimated in 2017 [31] that a sample count of at least 1000 per day justified the installation of total lab automation systems and when are other pressures prevail such as unavailability of staff or the need to reduce turnaround time, task-targeted automation can be sensible for 500 samples per day. Total lab automation requires a long-term commitment. As an example, the installation of a total laboratory automation system described by Lou et al. [35] required almost 2.5 years to complete.

Figure 3: Examples for different degrees of automation in a bioanalytical / bioprocessing laboratory. Devices in different scales. Exemplary vendors in brackets.

In the “middle of the automation spectrum”, there is a broad range of specialized or flexible automated or semi-automated workstations and instruments. A semi-automated procedure is a method in which some steps are automated while some steps are carried out by a human operator. Advantages are greater flexibility compared to full automation and reduced operation time compared to manual work. Compared to fully automated systems, however, more operator time is required in semi-automated methods and operators have more influence on the results. The suitability of a system in a specific case is determined largely by the sample volume and economic considerations. The process(es) to be automated and the degree of flexibility required also play roles in the decision of the most suitable automation strategy. For example, Lou et al. [35] reported that for large sample volumes (about 4 million tests per year) in the clinical laboratory, their multi-disciplinary total laboratory automation system turned out to be a great benefit in analysis time reduction, but for “small” sample volumes, a semi-automated workflow including manual centrifugation and sample loading on the analyzer was more efficient.

The implementation of automated systems or methods requires the transfer of established manual methods to the automated workstation. Ideally, automated methods can be developed quickly. A fast establishment of methods requires, among others, that the control software is intuitive to use for laboratory scientists and technicians.

1.6. Online sensors and statistical methods for real-time monitoring

For semi-automated real-time monitoring of chromatography processes, commercial non-destructive online sensors were used in combination with statistical models. Optical and electroanalytical sensors were used.

The electroanalytical sensors were a pH probe and a conductometer. These are standard online sensors in chromatographic workstations due to the importance of their respective signals and due to their robustness. A pH probe measures the potential between two electrodes, which is based on a chemical equilibrium [36]. The potential is related to the activity of H_3O^+ in the solution which is related to the H_3O^+ concentration at a certain temperature. A conductometer or conductivity sensor measures the resistance of an alternating current in a solution between two electrodes and returns a conductivity value in mS/cm. The conductivity is also temperature dependent due to varying mobility of ions in the solution.

Optical sensors use electromagnetic radiation and filters for wavelength selection, to obtain information about the sample. For aqueous samples, wavelengths between about 200 nm (UV) and 25 µm (mid-IR) are interesting. In optical sensors, several effects can contribute to the measurement signal: absorption of radiation, absorption with emission of a lower frequency (fluorescence), scattering, refraction, diffraction, and reflection. Optical sensors often measure the differential signal between the sample and a reference chamber which can be filled with a liquid which represents the sample background. This setup increases specificity by reducing the signal contribution by the sample buffer, for example.

UV/VIS absorption is one of the most widely used optical methods because it is sensitive, robust and allows very fast detection. We used a three-wavelength UV/VIS detector alternating between 214, 260 and 280 nm.

Sensors for UV/VIS absorption, conductivity, pH and pressure are commonly used in preparative chromatography from benchtop to manufacturing scales. These sensors are robust and can be used to monitor process parameters such as conductivity and pH of buffers and feed solutions or breakthrough due to column saturation (UV/VIS signal). To obtain information on different species which are present in the feed solution, more distinguished analyses are needed. Sensors and associated data evaluation methods for rapid bioprocess monitoring have been reviewed in numerous articles, e.g. [7], [15], [37]–[41]. Additional to the aforementioned sensors, we used a multi-wavelength fluorescence detector, a multi-angle static light scattering (MALS) detector, an infrared (IR) detector and a refractive index (RI) detector. The arrangement of sensors is shown in Figure 4. The sensors will be briefly described. More details can be found in [19], [23], [42].

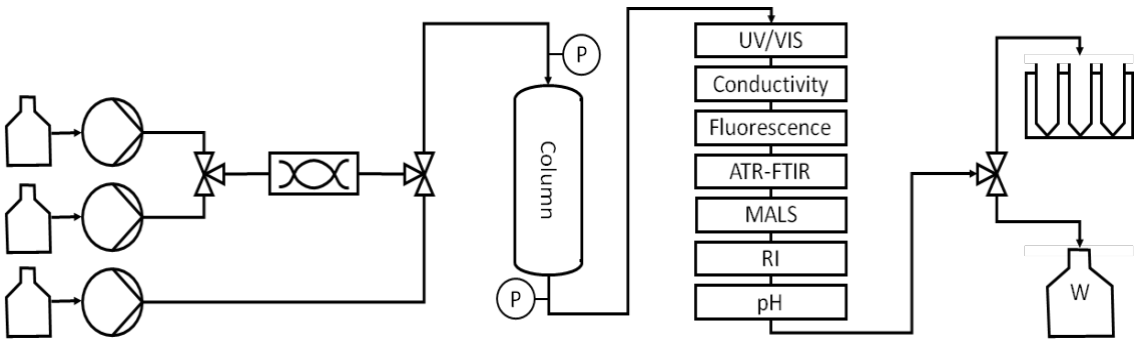


Figure 4: Schematic representation of the online monitoring system with solvent reservoirs, feed container, pumps, inline mixer, chromatography column, pressure sensors, online sensors, fraction containers and waste container; reproduced from Walch et al. 2019 [23]**.**

Refractive index and light scattering sensors measure interaction with the sum of sample components. Static light scattering, as in the used MALS detector, measures the relative amount of scattered light. It can provide information on the particle sizes in a sample [43] because larger particles scatter more light and in wider angles. Therefore, MALS detection is often used after size exclusion chromatography where components are separated based on their (hydrodynamic) size to estimate their molar masses.

The refractive index of a solution represents the degree to which light is refracted by the sample and, as mentioned before, is unspecific to the sample components. The refractive index is also influenced by the changing buffer used to elute the product, as can be seen in Figure 2 of Publication II [42].

More complex optical methods are spectroscopic sensors which record signals over a range of wavelengths using optical filters. Spectroscopic analytical methods are very useful because the interaction of electromagnetic radiation and chemical structures such as covalent bonds, e.g. amide bond or aromatic rings, give specific signals which can be used for structural characterization or quantification.

Mid- and near-infrared (MIR and NIR) absorption spectroscopy are popular analytical methods in protein analysis due to the specific absorption spectra, also referred to as IR “fingerprints”. The absorption of IR radiation causes molecular vibrations and the intensity of absorption at different wavelengths correlates to the presence of certain molecular structures of the protein or other analyte [44]. IR spectroscopy can be used for quantification if the absorption spectrum is known, i.e. by calibration, as well as for structural identification by comparison of the recorded spectrum with spectra of other (similar) substances. Suitable MIR technology was used for monitoring of bacterial cell integrity [45] and for quantification of *intracellular* product concentration during fermentations [46]. We used a detector based on attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR) which allows fast measurements and a high signal-to-noise ratio. A drawback of IR is the strong and temperature-dependent water absorption. In our data, the influence of water and buffer absorption was reduced by subtracting reference spectra of (1) deionized water recorded before analysis and (2) a spectrum of the buffer during the run just prior to elution. The first correction was done by the instrument software OPUS (Bruker), the second during data preprocessing in the statistics software R [47].

Intrinsic fluorescence of proteins derives from the aromatic amino acids tryptophane, tyrosine and phenylalanine and is influenced by the surrounding of these residues in the protein and by the sample matrix. Fluorescence emission spectra can give information about the correct folding of proteins and might allow distinction between different sample species due to different intensities at certain excitation and emission wavelengths. The detector we used recorded emission spectra from 236 to 795 nm in a resolution of 0.3 nm after excitation at six different wavelengths (265 nm, 280 nm, 289 nm, 300 nm, and 400 nm) and either one of two filter widths (10 or 40 nm). A reference channel was indicating the intensity of the xenon lamp and the integrity of the optical fibers. Due to the switching between channels which correspond to different excitation wavelengths, each fluorescence variable was available every 16 seconds.

Measurements of UV/VIS, RI, conductivity, pH and MALS detectors were available every second because no spectra over a range of wavelengths were recorded as for fluorescence and IR sensors. For all sensors, we chose setups with as many variables as possible (wavelengths, angles) to obtain as much information as possible. Multivariate data analysis methods were used to select the variables which are (most) useful to estimate concentrations of active proteins and impurities. In statistical terms, real-time estimation is called “prediction”

because attributes are estimated (from online sensor data) before they are determined (by offline analytics). Both terms, *real-time estimate* and *prediction*, are used in this thesis.

The *variables* on which a prediction is based, are the *predictors*. For example, the intensity of UV absorption at 280 nm can be a predictor of protein concentration, which is the *response* of the model. The most useful variables to predict a response are those which change with varying response (in this case concentrations), i.e. they have a covariance. The useful covariant predictors can be distinguished from “useless” variables by principal component analysis (PCA). For the type of data set obtained, partial least squares (PLS) regression (also called *projection to latent structures*) is a popular method. One step in this multi-step regression method is PCA which reduces the dimensionality of the data set significantly. PLS regression models formulate the response as a linear combination of a specified number of so-called *latent variables* which are derived from the original variables.

Cross-validation was used to avoid model overfitting which is the undesired modeling of the noise in the data. In cross-validation, the data set is split into subsets which are then used to train and test the model, respectively. The test data is used to calculate an observed error score, for example the root mean squared error (RMSE) or the mean relative deviation (MRD). For example, two chromatography runs can be reserved for model testing while the remaining runs are used for model training. This procedure can be repeated automatically until all data has been used for testing once. The average of the generated test errors is returned as the cross-validated error.

PLS, as a linear modeling technique, has at least one advantage and one disadvantage in the present case. The disadvantage is that linear methods best represent linear relationships, such as the concentration of dilute solutions and their UV/VIS absorbance. Linear modeling techniques are not as well suited for non-linear relationships as they are present in chromatography of complex solutions and the saturated online sensor signals [48]. The advantage is that linear methods can be trained faster than non-linear methods due to their relative simplicity. Large computation power and time were needed to optimize non-linear STAR models [19], especially when combined with different preprocessing methods and settings which are, however, indispensable for spectroscopic data [49]. A rational approach to find the optimal settings for data preprocessing was for example reported by Feidl et al. using a tree-based method [50]. Pais et al. used a genetic algorithm-based approach to find optimal preprocessing methods for fluorescence online data of an upstream process [51].

The developed models were solely data-based. Other common types of process models are mechanistic or hybrid models. The latter use combinations of mechanistic and data-based model parts which allows interesting advantages over the respective single-sourced model types [52], [53]. All models need some form of process data for calibration and validation. Besides the technique or model structure, the quality and quantity of the data used for training and testing are the basis for model performance [54]. It should be noted that statistical models are no virtual representations of the unit operation processes [12] and can therefore not be used for *digital twin*-assisted product applications but for process monitoring and as part of a process control strategy.

2. Objectives

The objectives of the thesis were (1) to establish semi-automated methods for sample preparation for determination of product purity and activity; and (2) to transfer the real-time monitoring system developed in-house [19], [23] to two industrial partners' sites.

The objectives concerned analytical methods of very different kinds. Biochemical analytical methods for quantification of impurities and binding affinity were transferred to an automated liquid handling equipment and established for routine analysis. The aims were to accelerate data generation and to increase standardization to achieve a reduction of variability in the generated data. This was required because a large amount of data was needed to train data-based prediction models for real-time process monitoring. During the development until transfer to the new sites, data was collected over a time span of 5 years by several analysts. Standardization by semi-automation was anticipated to increase reproducibility as was reported elsewhere [55]. Therefore, I hypothesized that semi-automated analytics is more precise and less time consuming than manual analysis.

The second objective was the transfer of a real-time monitoring system for chromatographic separation of proteins to two industrial partners' sites. The hypothesis was that a transfer of the models is possible because equivalent or similar equipment and materials were used at all sites. The transferability of the models should be evaluated in terms of their prediction errors on test runs performed at the new sites. A direct transfer of models developed at one site to another would save time and resources for data collection and model training.

3. Results

The results generated in this thesis were published in two publications as first or shared-first author. During the thesis, contributions to another publication were made. Author contributions to the publications can be found in Section 8.

3.1 Publications as first or shared first author

Publication I: Semi-automation of process analytics reduces operator effect

Bioprocess and Biosystems Engineering (2020) 43:753-764; Springer.

Anna Christler, Edit Felföldi, Magdalena Mosor, Dominik Sauer, Nicole Walch, Astrid Dürauer, Alois Jungbauer

https://doi.org/10.1007/s00449-019-02254-y

Publication II: Technology transfer of a monitoring system to predict product quantity and purity of biopharmaceuticals in real-time during chromatographic separation

Biotechnology & Bioengineering (2021) 1-12; Wiley.

Anna Christler*, Theresa Scharl-Hirsch*, Dominik Sauer, Johannes Köppl, Cabir Toy, Michael Melcher, Alois Jungbauer, Astrid Dürauer

*equally contributing authors

https://doi.org/10.1002/bit.27870

3.2 Other publications

A two-step process for capture and purification of human basic fibroblast growth factor from *E.coli* homogenate: Yield versus endotoxin clearance

Protein Expression and Purification (2019) 153, 70-82; Elsevier.

Dominik Georg Sauer, Magdalena Mosor, Anna-Carina Frank, Florian Weiß, **Anna Christler**, Nicole Walch, Alois Jungbauer, Astrid Dürauer

https://doi.org/10.1016/j.pep.2018.08.009

3.3 Summary of results

Laborious sample preparation steps in the protocols of biochemical analytics for the quantification of process-related impurities and binding affinities of biopharmaceuticals were semi-automated, as described in publication I [24]. The relative standard deviation of results among four different operators using the semi-automated method for DNA determination over a time span of 6 months was 48% lower compared to manual (7.1% vs. 13.7%). In this assay, the median of results was 17% higher with the semi-automated method compared to the manual results. Precision of ELISA results and the absolute values were similar when one analyst performed all tests (10.9% vs. 8.3% manual). Thus, semi-automation of offline analytics did not increase the precision of results compared to an experienced analyst. However, operator effects could be reduced in one assay. Operator hands-on-time of a polishing chromatography run with 15 fractions was 5.8 hours with the manual methods. With the support of the liquid handling station, only about 3.4 hours of manual work were required. Thus, hands-on time was reduced by about 2.4 hours or 41%. Even more time could be saved for early purification steps where 10 – 27 times higher dilution factors were necessary.

The developed real-time monitoring system was transferred to two other sites, as described in publication II [42]. Equivalent systems were installed at the new sites. Sensor signals of the online monitored chromatography runs were similar between the sites except for the fluorescence sensor which was an in-house assembled prototype. Due to the differences in the fluorescence data, the developed prediction models were not directly transferable to the new sites. Exclusion of five excitation wavelengths in the range of 300 to 400 nm and the corresponding emission data, and adjustment of data preprocessing was necessary. After the adjustment of model predictors and data preprocessing, prediction errors were on average twice as high (per model 0.9 – 5.7 times) at the new sites compared to the training site. One reason for the higher errors was the limited sensor robustness of the fluorescence detector. Also, for some impurities, higher or lower concentrations were measured at the new sites compared to the training site which led to model extrapolation. Poorer performance of data-based models outside their design space was expected. At all sites, prediction of purity parameters led to higher prediction errors compared to the prediction of product concentration. Model training for new processes and products at the new sites allowed prediction of six quality attributes with mean relative deviations between 1 and 33%. Among the quality attributes were product charge variants which are structurally very similar to the POI.

4. Discussion

On the way to industry 4.0 and even 5.0[4], automated processes and process control strategies are required in biopharmaceutical manufacturing [1], [2]. As biopharmaceutical products become more diverse, manufacturing processes must be developed quickly, be flexible and controllable. Biosimilars and biobetters on the market increase the economic pressure on manufacturing companies. Real-time process monitoring enables a small step towards a modern digital factory and real-time batch release.

4.1. Semi-automated offline analytics

Data generation for model training represented the bottleneck during development of the real-time monitoring system. Accurate representative process data was needed for model calibration and testing. The maximum sample count per week was: 4 chromatography runs (two of each POI) with 18 fractions each and 6 analyses per sample amounting to 432 analyses. Counting only the laborious biochemical analyses (dsDNA, HCP, endotoxins, and binding affinity), the total count of analyses was 252 (Table 1). For these analyses, two full-time equivalent analysts were occupied.

Table 1: Maximum numbers of samples and biochemical analyses for model training per week.

	FGF-2	**IgG**
Chromatography runs per week	2	2
Average number of samples per run	18	18
Biochemical analyses per sample	4	3
Number of samples for biochemical analysis per week	144	108
Total	252	

The directly product-related quality attributes – protein concentration and contents of monomer, HMWI, LMWI and charge variants – were quantified by HPLC or UPLC and required only minimal sample preparation: dilution with phosphate buffers, filtration, filling in HPLC/UPLC vials and capping. Special automated systems for sample preparation for HPLC/UPLC are available, for example accroma®[5], but most liquid handling systems are not designed to filter samples and cap HPLC/UPLC vials. In our case, relatively small dilution factors in the range of 0 – 20 were needed for most HPLC/UPLC analyses. For size exclusion chromatography of FGF-2, factors of 5 – 100 were needed. Thus, sample preparation for HPLC and UPLC measurements did not need extensive operator time and no special additional equipment was justified for the respective sample preparation tasks. In contrast, large dilution factors and thus extensive manual pipetting was required for the analysis of DNA, HCP, endotoxins, and binding affinity, especially in samples from early-stage purification steps. Analyte concentrations in capture stage purification and analytical ranges of assays are shown in Table 2 and 3.

[4] Industry 5.0: While the idea of Industry 4.0 is a completely smart and automated way of production, Industry 5.0 builds upon the achievements of Industry 4.0 but aims to be more human-centered and sustainable. See for example: https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en, accessed August 31, 2021.

[5] https://accroma.com/, accessed October 07, 2021.

In each dilution step, inevitably a small error is introduced. When pipetting viscous concentrated samples, the pipetting technique can have a large influence. From own observations, the most common errors in pipetting are a too high speed and an inconvenient angle of holding the pipette. Several aspects should be considered for pipetting accuracy (compare the infographic in Section 9.1). Many, if not all of them, reduce the speed without immediate obvious benefit. Therefore, repeated training of analysts is necessary. A good pipetting technique is always required but is especially important when pipetting viscous biological samples. Automation of sample dilution for binding affinity (Table 3) eliminated a lot of manual adjustment of pipettes because the volumes of sample and buffer had to be adjusted to each fractions' protein concentration.

Table 2: Ranges of analytes in process samples from chromatographic capture, analytical ranges of assays and analyte concentrations relative to upper limits of quantification (ULOQ). Adapted from Christler et al. 2020 [24]**.**

Protein of interest (POI)	Sample component	Concentration range in fractions[1]		Analytical range of assay (mL^{-1})	Analyte concentration in % of ULOQ
		per mL	in ppm of the POI		
FGF-2	FGF-2 concentration	0.01 – 42 mg	–	–	–
	dsDNA	50 – 2,000 ng	8 – 5800	3.91 – 500 ng	10 – 400%
	HCP	20 – 500 ng	10 – 4670	0.39 – 25 ng	80 – 2,000%
	Endotoxins	20 – 188,000 EU		0.01 – 5 EU	400 – 3,760,000%
IgG	IgG	0.004 – 32 mg	–	–	–
	dsDNA	≤ 1.95* – 100 ng	0 – 900	1.95 – 250 ng	≤ 40%
	HCP	0.060 – 30 µg	2,200 – 10^6	2.11 – 135 ng	44 – 22,222%

[1] Distributions over the fractions can be found in [18] (Figures 1-3, 5 and 6) for FGF-2 samples and in [23] (Figure 2) for IgG samples.

* Below the lower limit of quantification (LLOQ).

Table 3: Samples and dilutions for binding affinity assay using surface plasmon resonance on a Biacore instrument.

POI	Concentration of POI (mg/mL)	Concentrations required for assay	Dilution factors required[6]
FGF-2	0.01 – 42	1 – 20 nmol/L = 1.7×10^{-5} – 3.4×10^{-4} mg/ml	29 – 2,470,588
IgG	0.014 – 32	10 – 100 nmol/L = 0.0014 – 0.014 mg/ml	0 – 22,191

[6] The minimal dilution factor (DF) for FGF-2 samples (29) represents a sample of 0.01 mg/ml that is diluted to 20 nM (the highest concentration to be measured). The maximal DF for FGF-2 samples (2,470,588) represents a sample of 42.0 mg/ml FGF-2 that is diluted to 1.0 nM (the lowest concentration to be measured). Similarly, for IgG: the minimum required sample concentration was 100 nM or 0.0144 mg/ml which has a DF of 0. A sample of 32.0 mg/ml must be diluted 22,191-fold to reach 1.0 nM.

Relatively larger volumes and smaller dilution factors were favored over relatively small sample volumes and large dilution factors since the relative pipetting error decreases with larger volumes. Lippi et al. [56] observed an approximately 7-fold reduction in the coefficient of variation for both intra- and inter-individual pipetting precision of 20 operators when pipetting 100 µl of water (!) compared to 10 µl and another 3- to 4-fold reduction from 100 µl to 1000 µl. Probably from extrapolation of the findings, they conclude that "manual pipetting of 1 µL may be considered *unacceptable* and manual pipetting of 10 µL may be acceptable in some conditions but not in others." [56]. These numbers show that care must be taken when the pipetting of small volumes cannot be avoided such as in ribonucleic acid research, for example. Similarly, dilution factors should only be as large as necessary to be in the range of quantification (compare Table 2) and exclude or reduce interfering effects from other sample components to a minimum.

Due to the introduction of the liquid handling system during the development of the real-time monitoring system, data from manual assay execution (from four analysts) and also from the semi-automated methods was used for model training. Therefore, data from manual and semi-automated methods must be equivalent to avoid biasing the models. Some differences between manual and semi-automated methods were unavoidable. For example, in some steps, larger volumes had to be used with the liquid handling station because less pipetting tools were available. With the semi-automated methods for sample preparation for binding affinity, HCP, CHO cell dsDNA and endotoxins, equivalent results as with manual methods were obtained. In one assay, *E. coli* host cell DNA determination, systematically higher results (median: +17%) were obtained. Higher results might have been caused by smaller dilution factors because a dilutional bias was observed (see publication I [24], Figure S4 in Supplementary Material). The definitive root cause of the differing results between manual and automated methods was not identified. Since the confidence intervals overlapped due to the relatively large variability of the assay results, the semi-automated method was used for routine analyses. No obvious influence on the model calibration was observed in the models' predictions.

Automated plate washing needed significantly more time than manual and thus was not useful in routine analysis. Similarly, Tornel et al. described the increased turnaround time (= longer analysis time) due to the automation of an additional step (centrifugation) with the benefit of reduced manual work [57]. They accepted the increased turnaround time due to overall increased laboratory efficiency. Van de Bilt et al. [27] reported that three washing cycles were needed in their automated plate washing procedure compared to one in the manual method. Nevertheless, the automated method was twice as fast overall due to parallel preparation of 144 samples in 5 dilutions each. We could process at most two 96-well plates containing each 14 samples in 6 dilutions (i.e., in total 28 samples) plus reference standards and blank. In this case, manual plate washing was more efficient. Even if the number of dilutions was reduced to 5 or four (see publication I [24], Figure 4), the maximum number of samples in one run would be limited to 16 or 20 per plate, i.e. 32 or 40 in total – far away from 144. The comparison of the three publications shows that it depends on the circumstances whether the automation of a step in a workflow is sensible or not. It depends for example on the surrounding work steps, the possible degree of parallelization, the availability of operators and the time pressure.

Semi-automation of biochemical analyses reduced manual repetitive steps, mainly pipetting, in our analytical workflows. This reduced the dependency of analytical results on operators. Our data confirm other sources which stated that results generated by a liquid handling system were not more precise than those of trained and experienced analysts [55]. Yet, technical solutions eliminate random human errors. It would have been interesting to evaluate whether the implementation of the automated methods lead to reduced re-testing of samples due to the decreased chance of random human errors.

Automated systems are also “stupid” in the sense that they just execute programs. Machines lack creativity and common logic sense [33]. If not equipped with respective sensors, they ignore possible complications such as air bubbles, foam, missing samples, mismatching volumes, etc. Appropriate settings for the speed of aspiration and dispense were critical for protein solutions since they tend to form bubbles and stick to surfaces. As described in publication I [24], the pipetting accuracy of non-protein solutions was better than that of protein solutions. Reverse pipetting helped to reduce bubble formation in the ELISA sample buffer which contained 1% BSA (i.e., 10 mg/ml) and 0.05% detergent (Tween 20). However, the standard mode of aspiration and dispense (“forward” pipetting) was more accurate. Advanced machine vision technologies are advisable to increase confidence when pipetting such solutions [31].

A side-effect of automation was the need for other skills in the lab: “digital” knowledge to operate automated equipment. In an automated laboratory, different training of personnel or even different personnel is needed compared to a classical wet lab. New technologies often bring uncertainty with them. Some people fear losing control over workflows, others fear to become redundant by new technologies [33]. In our lab, analysts received the introduction of the liquid handling station differently. For some analysts, the lower precision was a reason they felt uncomfortable using the semi-automated methods. Others welcomed the reduction of manual repetitive work. Continuous training was necessary to make people understand and use the benefits of the technology effectively. The skepticism of the beginning changed to acceptance when the users became more familiar with the system and the advantages became more obvious. Besides sample dilution, automation of a sequence of repetitive time-critical steps, such as starting and stopping the staining reaction of an ELISA plate, were the most useful automation tasks.

Another advantage of automated liquid handling systems can be the simplification of assay optimization based on a Design of Experiments (DoE) approach which can be tedious to plan and execute manually. The same is true for method development and validation where sets of samples must be created such as series of standard additions. The system control software would need to offer these functionalities and should assist in data evaluation and reporting.

Automated systems become more efficient when they are connected and communicate with each other [58], for example a liquid handler and a spectrophotometer. Information about erroneous samples (e.g., due to mismatching volumes or bubble formation) could be transferred to the protocol of the analytic device [59]. Analysts would need to check only one report. In a highly digitized lab, the information could even go both ways, i.e., from the analytical instrument back to the liquid handler suggesting the analyst and instrument which samples should be repeated, for example because the dilution needs to be adjusted based on the measurements.

Further automation and/or acceleration of sample analysis could be achieved by assay miniaturization, for example by “lab on a chip” or microfluidic devices [60], [61]. Sample and reagent volumes could potentially be reduced, as well as costs and the amount of waste. *Gyros Protein Technologies*[7] offers automated immunoassays at the nanoliter-scale. Basically, every ligand-receptor pair can be used for specific quantification and binding affinity evaluation [62], [63]. Compared to ELISA, results can be obtained faster, with less manual work, requiring fewer equipment and less reagents[8]. Although not fully miniaturized, a similar platform for endotoxin

[7] https://www.gyrosproteintechnologies.com/gyrolab-technology; accessed 30 August 2021.
[8] https://www.gyrosproteintechnologies.com/gyrosizing-your-elisa; accessed 30 August 2021.

determination is the Sievers Eclipse BET marketed by SUEZ Water Technologies & Solutions[9]. Microfluidic DNA quantification is proposed in research and development, mainly for low concentration clinical applications [64] but also in the higher concentration range of 10 – 110 µg/ml [65]. The latter is a quite narrow range of quantification compared to the PicoGreen dying method we used [24] which quantified dsDNA from about 4 to 500 ng/ml.

Another relevant application of microfluidics in bioprocessing that should be mentioned in this context is the scale-down of manufacturing unit operations to accelerate investigations or process development or optimization tasks, e.g. [66], [67].

Pros and cons of conventional manual and (semi-)automated bioanalytical workflows are summarized in Table 4.

Table 4: Possible advantages and disadvantages of conventional manual processes and (semi-)automated processes in bioanalytical labs.

	Manual workflow	(Semi-)Automated workflow
Advantages	• Usually well-known workflows • Experience and equipment often available • Full flexibilty	• Parallelization of workflows possible leading to time reduction • Less manual work, therefore either time saving or easier or safer workflow • Higher degree of miniaturization possible (nanoliters in microfluidic techniques) • Traceability due to electronic documentation (audit trail, barcodes, RFID) • Possibility of automated data evaluation and reporting
Disadvantages	• Repetitive tasks are error-prone and can lead to fatigue • Traceability depends on manual documentation	• Restriction of possible assay “configuration” due to dependency on software and machines • Digital skills needed in addition to chemical and analytical skills • Less or no experience available in many labs

4.2. Real-time monitoring of biochromatography

Many new technologies or systems are being developed using model solutions like synthetic distinct mixtures of few components, for example the protein of interest (POI), variants of it and/or typical impurities, e.g. [68], [69]. In our lab, we took the challenge to develop a system using two industry-relevant purification processes and feed materials to test and show usability in an industrial-like setting. Feed materials from fermentation and cell culture, as described in Section 1.3, were used. Fermentation and cell culture supernatants are complex biological mixtures containing the POI, different product variants, cell debris, media components, and potentially other contaminants such as viruses. In the load materials, many different impurities were present in relatively high concentrations (see Table 2). The POI and most of the impurities have very different properties regarding hydrophobicity, charge and/or size which is the basis for their separation in downstream processing. However, this complex mix complicated the development of data-based prediction models for real-time monitoring in several ways. First, some offline analyses were influenced by the sample matrix, most dominantly quantification of

[9] https://www.suezwatertechnologies.com/products/analyzers-instruments/sievers-eclipse; accessed 30 August 2021.

low concentrations of dsDNA in IgG samples by PicoGreen staining. The quality of offline measurements influences the prediction quality by the statistical models because the model assumes the reference to be the *true value*.

Secondly, the high concentrations of proteins and other impurities caused fouling in the monitoring system. Fouling, which is the attachment of proteins and other species onto surfaces, reduces the efficiency of unit operations such as filtration or chromatography by blocking pores and/or ligands. Some proteins, especially membrane proteins easily adhere to surfaces due to their amphiphilic nature. Fouling is a common problem in bioprocessing. PAT applications have been applied to monitor fouling in chromatographic columns [70]. In the manufacturing equipment, especially in pipes, fouling reduces the available cross-section. According to the continuity equation (conservation of mass), when the flow is kept constant (by the pump), the flow velocity increases linearly proportional to the reduced cross-section. Thus, the pressure drop in the tube system increases due to friction, according to the Darcy-Weisbach equation with the square of the flow velocity. Some instruments' flow cells are sensitive to pressure due to their delicate architecture made for high precision. For this reason, sensors of the developed real-time monitoring system were arranged in a sequence accounting for the pressure sensitivity (high to low pressure tolerance) (compare Figure 4). Fouling can even lead to clogging of narrow channels or pipes which would stop the process. To avoid this, equipment was rigorously cleaned with acid or caustic solutions between batches. MALS and RI detectors were also sensitive to extreme pH values and had to be cleaned otherwise. In an industrial setting, increasing the robustness of the sensors would be needed to allow for a closed process, including cleaning procedures, as far as possible.

The third way in which the authentic load materials complicated the development of prediction models was the saturation of sensors. The complex biological solutions posed a challenge for the sensitive analytical instruments. For quantitative analysis, the relationship between analyte concentration and signal intensity should ideally be linear for best sensitivity and accuracy. This is only possible in the low concentration range. Exact numbers depend on the method and analyte but are often below 1 mg/ml (e.g. BSA fluorescence at 280 nm [excitation] / 340 nm [emission] [71]). Sensor saturation is especially problematic at the borders of the peak where pooling decisions must be made. In early-stage purification, saturation of sensors is especially an issue due to the high impurity content. In later downstream unit operations, sensor saturation will be less pronounced in the concerned regions of the chromatogram because the impurity concentrations are much smaller. A variable pathlength UV/VIS spectrometer was applied by Brestrich et al. [69] to extend the dynamic range and selectively quantify and pool two-component protein mixtures of up to 80 g/L after chromatographic separation, however with a long measurement cycle time of 30 seconds.

Protein activity could not be predicted with the statistical models. The offline measured dissociation constant was equal within the range of the method's variability in all fractions (1 – 4 nM for FGF-2 [18] and about 20 nM for IgG binding to TNFα[10]). No inactive POI was detected by the binding assays. Thus, it was impossible to correlate the sensor signals to the offline determined quality attribute. Inactivated POI would need to be spiked to the active in some of the fractions. However, we wanted to remain with the real processes.

[10] Note: The dissociation constant for the IgG (Adalimumab) is no generally valid number as it depends on assay and instrumental setups. Lower values, corresponding to stronger binding, have been reported in literature.

4.3 Transfer of the monitoring system to two new sites

The developed real-time online monitoring system was transferred to the industrial partners' sites. The transfer included the following steps:

1) Installation of the sensors and database at the new sites;
2) Transfer of process materials (columns, load material) and purification protocol;
3) Training of operators on the handling of the monitoring system;
4) Performance of online monitored chromatographic runs;
5) Comparison of sensor signals and model predictions from new sites with the training site.

Predictive models for five quality attributes were trained using the data from the training site based on PLS regression, as described in section 1.6, and applied to the new sites data [42]. Measured dsDNA concentrations were very high at site A and very low at site B (Figure 5). Thus, the model for dsDNA prediction was extrapolated to partially unknown concentrations which failed as shown by large deviations of the predictions from measured values. Generally, the models for POI concentration (quantity) and monomer content predicted reasonably well, but the models for HMWI content, HCP and dsDNA were not really acceptable as reflected in the high RMSEs compared to the Null models (Table 1 in [42]). They would need to be retrained with data from the respective sites. Thus, the hypothesis of direct transferability of the models could not be accepted under the circumstances. If all sensors were robust and the results of the processes more similar, a direct application of the models might be possible.

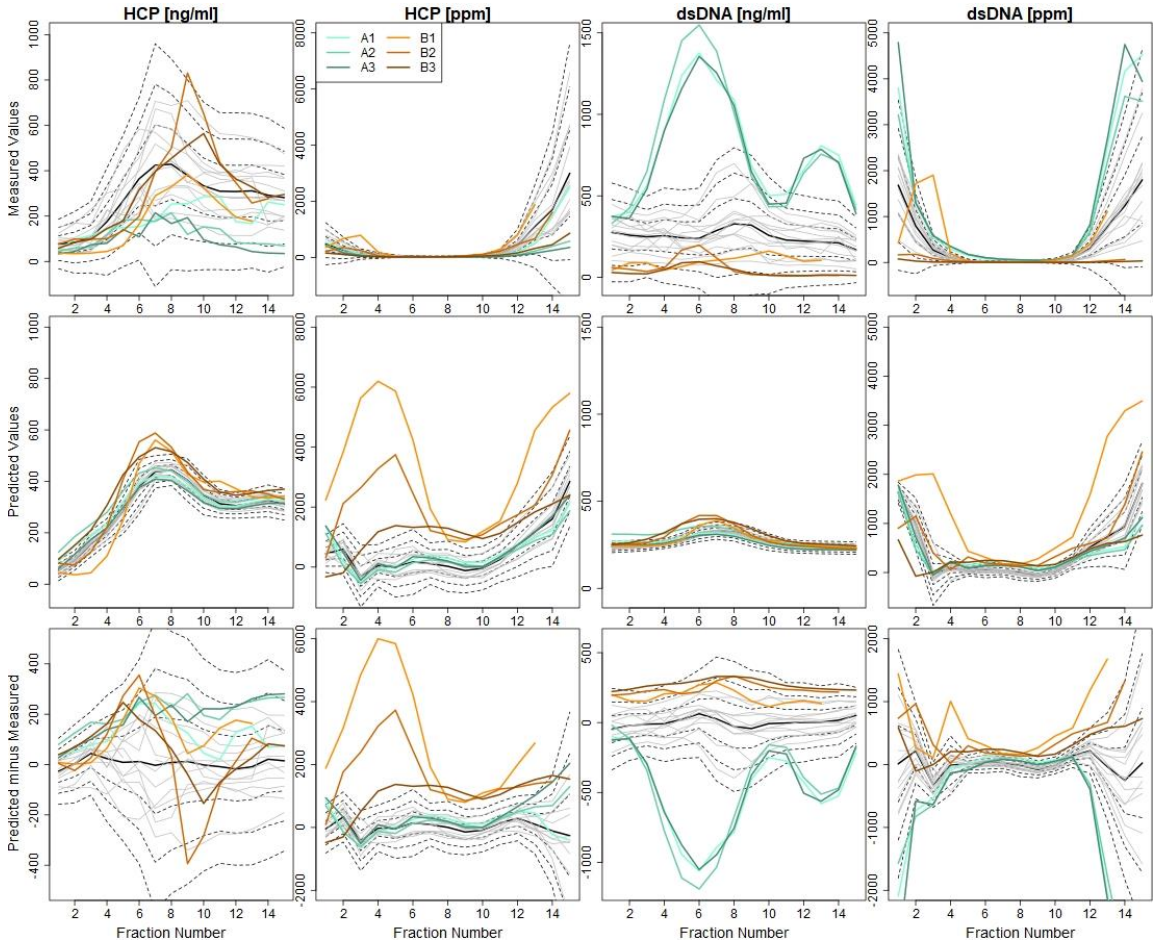


Figure 5: Comparison of responses HCP and dsDNA modeled and predicted either in units ng/ml or in ppm.

The sensors are the critical point of the monitoring system. They are expected to deliver highly sensitive measurements but need to be robust at the same time. In our study, the compact commercial sensors (compare Figure 6a) were robust and delivered reproducible signals at different sites [42]. The in-house assembled fluorescence sensor, a prototype (see Figure 6b), was not robust. It gave different signal profiles and intensities at different sites. Even the reference channel showed very different signals and behaviors at the three sites. As described in publication II [42], smoothing, taking the first derivative and normalization to equal length was useful for excitation at 260 and 280 nm. At higher excitation wavelengths, the signals remained very different at one site. Furthermore, the relatively fast signal decrease over time due to aging of lamp and optical fibers was problematic. Close monitoring of signal intensity and data preprocessing was needed at all sites. A multiplexer is generally interesting because it allows the measurement of many different wavelengths in one device (see publication II [42], Figure 1). This set-up was chosen to screen as many excitation wavelengths as possible and obtain full emission spectra. One measurement cycle of all channels took 16 seconds and for this time, a constant value was used as predictor. In an industrial set-up, the number of fluorescence excitation wavelengths would be reduced and potentially also the emission wavelengths as it is commonly done in HPLC/UPLC fluorescence detectors. König et al. [71] used a miniaturized fluorescence detector without any moving parts for highest robustness for the prediction of biomass in upstream bioprocess using three excitation/emission wavelength pairs and scattered light at 850 nm. Fluorescence data can give very valuable information because intrinsic fluorescence of proteins depends on the secondary and tertiary structures. These structures, i.e. the correct protein folding, are essential for protein integrity and thus activity. Helgers et al. [72] achieved monitoring of mAb concentration in chromatography with UV/VIS and fluorescence data but not with ATR-FTIR or Raman spectroscopic data.

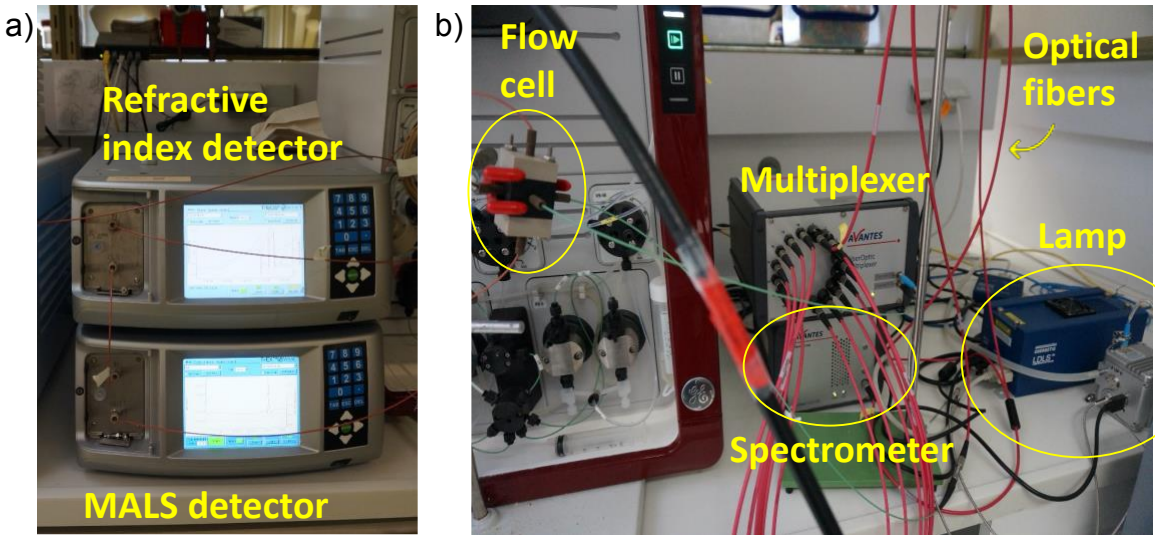


Figure 6: (a) The commercial refractive index and MALS detectors used in the study and (b) the fluorescence detector assembled in-house from several separate parts. Compare Figure 1 in publication II [42] **for a schematic representation of the parts of the fluorescence detector.**

Quality control measures were needed for all online sensors, e.g. light source intensity control, performance qualification test or measurement of a reference standard or spectrum before the sample(s). Absolute intensities varied from day to day, run to run or drifted over time. Data preprocessing was an important prerequisite to modeling and prediction. Kuhn states that “data preparation can make or break a model’s predictive ability” [13] which was confirmed in our case. Without preprocessing, IR and fluorescence data could not be used for predictive modeling. Transformations such as normalization and derivatization often help to remove or

reduce adverse effects such as scattering or baseline shift [49], [54]. Also “simple” methods such as removing variables based on expert knowledge or experience are effective. For the UV/VIS sensor, a reset to zero during system equilibration was sufficient. Sensors and the applied chemometric method including data preprocessing must also be able to detect and quantify analytes despite changing buffer backgrounds during elution [72].

In regression it is desirable to have a data set with more observations, i.e. replicate measurements (in this case: chromatography runs) than variables, i.e. different parallel measurements (in this case: sensors). In our data set, the number of variables (in the raw data about 17 830) was much larger than the maximum number of observations (246; however not all sensors were available in all 16 runs). Additionally, the variables were linearly correlated delivering redundant information. Multicollinearity occurs when predictors represent the same or similar chemical information. For example, similar wavelengths of UV/VIS or IR frequencies, especially the closer they are to each other, are being absorbed by the same molecular structures in the sample, only to a little different extent. Using both wavelengths in a model would not be helpful in predicting the response but increases the risk of model overfitting. Both issues, the unfavorable ratio of observations and variables and the multicollinearity of the predictors, were expected and were addressed by using PLS regression, a method particularly suited in such situations because it is based on data dimensionality reduction.

Simple models were preferred over complex ones. Whenever several model variants yielded equal or almost equal errors, the model with fewer predictors was selected. If a variable offers no benefit in the training set, it might worsen the prediction in a future case because it could represent the noise specific to the observations in the training data set. Another reason is practical: sensors might fail temporarily or permanently during a run, which they did sometimes in our case, and models which require fewer sensors have a higher statistical chance of stable performance. Lastly, model interpretation, if possible at all, becomes very difficult when models are complex. In contrast, Helgers et al. [72] pointed out that even if similar prediction qualities are obtained using a single sensor and a combination of sensors, respectively, the combination yields “different information, thereby providing potentially a more robust prediction” [72]. They therefore recommend using two orthogonal sensors, i.e. sensors which measure different aspects of a sample component or by a different principle, instead of only one.

Due to the narrow tubes and flow cells, the pressure drop over the sensors was about 1 bar. The flow rate was therefore limited to about 1.5 ml/min corresponding to about 57 cm/h or 7.2 and 8.2 minutes residence times for FGF-2 and IgG, respectively. Linear flow velocity in Protein A and IEX capture processes is commonly 100 – 400 cm/h and can be up to 1000 cm/h in capture processes [13]. Residence times are usually 2 – 5 min in capture processes for productivity reasons. To reduce the pressure drop, either less sensors in series need to be used or larger diameters of tubes and flow cells. However, larger flow cells would reduce the sensitivity of the sensors and potentially increase undesired back-mixing of the sample stream. A technical solution would be to split the flow stream in a bulk and a smaller stream for in-line measurements which comes at the cost of (1) an open system (risk of contamination) and (2) worse cleanability. Therefore, split-stream solutions are usually unpopular in hygienic production systems.

Besides the high pressure drop over the in-line sensor battery, another limitation of systems with this many sensors are the high investment costs. One aim was to evaluate which sensors are useful to predict which sample component. Unfortunately, this aim was not achieved. With one exception (monomer content), very different sensors were selected in the models (see Table 5). Different interactions between sample composition, signal preprocessing and modeling technique influence the choice of variables in the models.

Table 5: Comparison of sensors used in models for two different proteins and processes and modeling techniques in three cases. Colors highlight common features of the three publications. IEX: ion exchange chromatography. MALS: Light scattering at 43.6°, 90°, and 136.4°. PLS: partial least squares. Ppm: parts per million (of the POI). RI: refractive index. STAR: structured additive regression. UV: UV absorbance at 214 nm, 260 nm AND 280 nm.

Reference		**Sauer et al., 2019** [19]	**Christler, Scharl et al., 2021** [42]	**Walch et al., 2019** (sensors in bold increased model complexity but also prediction quality) [23]
Process and POI		FGF-2 purification with IEX chrom.	FGF-2 purification with IEX chrom.	IgG purification with Protein A chrom.
Regression / modeling method		STAR	PLS	PLS
Sensors used in models (units of the modelled response)	**POI conc. (mg/ml)**	UV + conductivity	UV + RI + conductivity	UV (**pH + RI**)
	Monomer (%)	*not evaluated*	UV + RI + fluorescence	UV (**RI + fluorescence**)
	HMWI (%)	*not evaluated*	UV + conductivity	UV + RI + (f**luorescence and IR** or pH and MALS)
	HCP	(ppm) UV + conductivity + fluorescence	(ng/ml) UV + conductivity + MALS	(ppm) UV + RI + MALS + (pH or **IR**)
	DNA	(ppm) UV + fluorescence + IR	(ng/ml) MALS	(ppm) UV + RI + MALS + (pH or **fluorescence**)

Another possible reason for the different predictors selected in the models are the different units in which the responses were regressed. The conversion from the measured response in ng/ml to ppm of the POI changes drastically the shape of the response and thus the relationship between response and variables, see Figure 5. The prediction of HCP at site B (orange/brown curves) in ng/ml is not accurate but acceptable but is completely wrong in ppm.

Data-based predictive modeling for real-time monitoring is a highly interdisciplinary task requiring expertise in downstream processing, analytics, automation, data science and statistics. Knowledge about the underlying process is required to identify faulty data which contains artifacts or results from equipment malfunction for example. Communication between disciplines must be established in the team and should not be underestimated. Similar to automation in the chemical laboratory, implementation of modern process control technologies such as PAT requires different skills compared to conventional methods: data mining, machine learning and artificial intelligence [4].

The developed real-time monitoring system can also be described as a semi-automated system because product was pooled manually based on model-predicted quality attributes. Similar to the situation of the liquid handling station and the plate reader discussed above, the real-time monitoring system will be more efficient when the single sensors were integrated into one system. The customized automation software XAMControl® (Evon, Austria) enabled coordinated data acquisition and correct alignment of the sensor data which would be cumbersome to do manually.

Another sensor that might be interesting for online monitoring is Raman spectroscopy. Raman scattering is molecule-specific and water does not interfere the measurement. However, the intrinsic weakness of the Raman signal makes it hard to be used in the relatively fast processes in downstream processing (compared to upstream processes). For example, Helgers et al. [72]

reported the successful monitoring of product concentrations in upstream processing and aqueous two-phase extraction by Raman spectroscopy. They used a 1.5-mW laser and only required a measurement time of 1 second per spectrum. They cumulated 3 spectra giving a total measurement time of 3 seconds. However, they could not use it for chromatography at all due to the large flow cell causing too much back-mixing of the effluent stream. Feidl et al. [50] developed a smaller flow cell to monitor mAb breakthrough after a protein A chromatography column by Raman spectroscopy. Even though a much more powerful laser (400 mW) was used, a measurement time of 1 minute was required. One minute is too long for automated peak cutting where pooling decisions must be made within seconds.

5. Conclusion

Both semi-automated systems investigated in this thesis, the liquid handling station for wet lab analysis and the real-time monitoring system for biochromatography, were not more precise than the corresponding manual methods. However, they substantially reduced operator hands-on-times and reduced operator influence. Although 17% higher results were obtained for one assay with the semi-automated method compared to the manual method, overall, the method transfer for four biochemical analytical assays to the liquid handling station was possible and successful. The transfer of statistical models from the training site to two other sites resulted in increased errors of prediction even though the systems and materials were equivalent. The models were thus not directly transferable.

Quality control and robustness of online spectroscopic sensors concerning integrity, function, fouling and aging is necessary for a robust operation of an online monitoring system. Real-time estimation of purity was the most challenging task and estimation of activity was not realized at all.

Semi-automation is recommended whenever repetitive steps represent a large part of the work and can be automated. Compared to full automation, more flexibility is allowed by semi-automated methods and investment costs for equipment are often lower. Different skills are needed when transitioning from conventional offline work to digital automated ways of analysis or manufacturing. For automation to be successful, there must be a good interconnection between human operators and machines. Usability, i.e. user-friendliness, of machines is key for efficiency of automated solutions. Lab automation systems need to allow intuitive workstyles for analysts. Analysts should be continuously trained and methods re-evaluated to maximize system efficiency.

6. Outlook

In a next step, an automated pooling algorithm would be added to the real-time monitored chromatography system to switch a valve routing the liquid stream after the sensors either into a waste or a product collection vessel. Walch et al. [23] and Sauer et al. [19] demonstrated exemplarily that it is possible to collect the product based on pre-defined criteria for each monitored variable with similar yields and qualities. Automated pooling can be realized by a simple algorithm switching a valve. Due to the noisy nature of the fluorescence sensor data for example, pooling decisions should be made only in near real-time to allow data interpolation (e.g. by a Savitzky-Golay filter [69]) and thereby assure that decisions are not made due to spikes to higher or lower predictions. Due to the physical distance between the sensors and the switching valve, the time delay for data acquisition, preprocessing, prediction and interpolation will be sufficient with our instrumental set-up.

The integration of all sensors into one compact unit would allow easier handling of the developed real-time monitoring system. Easy-to-use and robust hardware and software would need to be designed. A Design of Experiments (DoE) functionality integrated into the software would allow to control the relevant design space of the models and support system validation.

Another logical extension of the system is its application for process control for either continuous processing or critical unit operations, for example for difficult separations where disturbances easily influence the process performance. Measurements of the feed material (in a split-stream manner or in-line diluted to avoid sensor saturation) could be taken (at-line or in-line) before the column, concentrations estimated in real-time and process parameters adjusted accordingly. In batch-wise manufacturing, the benefits can be increased yield and/or productivity and/or quality [15] while continuous processing is only possible using online process control. For *robust* batch processes, the effort and investment is usually not justified [15].

One method to increase confidence of the predictions and thus of the pooling decision could be the addition of a mechanistic process model (Figure 7) [73], [74]. Mechanistic and hybrid models are of interest in research and industry in addition to data-based models because complimentary information can be obtained [53]: Statistical, data-based models deliver (quasi) real-time estimates for time points where data is available while the prediction of future events is only possible with mechanistic models. Another advantage of mechanistic models is the possibility to close mass balances. The process could be predicted by the mechanistic model based on the composition of the load material. Taking the feed volume into account, total amounts of loaded product and impurities could be calculated and used to correct real-time estimations and predictions. On the other hand, predictions by the mechanistic model could be updated by the statistical part using real-time estimates. Weights could also be given to the mechanistic and the data-based model parts in different stages of the process or for specific responses when one of the model parts is known to outperform the other [75]. A mechanistic model can also help to determine the largest possible design space for safe operation which allows greater flexibility in operation when filed to the authorities accordingly [9].

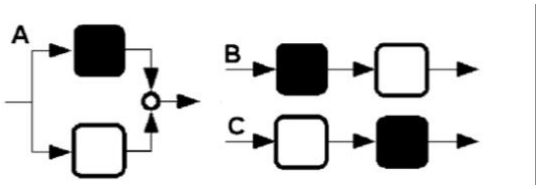


Figure 7: Schematic representation of hybdrid model structures: A, parallel, B and C, serial structures. Black boxes represent data-based models or model parts, white boxes represent mechanistic models or model parts [75]**.**

7. References

[1] M. Sarkis, A. Bernardi, N. Shah, and M. M. Papathanasiou, “Emerging Challenges and Opportunities in Pharmaceutical Manufacturing and Distribution,” *Processes*, vol. 9, no. 457, 2021.

[2] M. M. Papathanasiou and C. Kontoravdi, “Engineering challenges in therapeutic protein product and process design,” *Curr. Opin. Chem. Eng.*, vol. 27, pp. 81–88, 2020.

[3] EMA, “ICH Guideline Q8 (R2) on pharmaceutical development,” 2017.

[4] A. L. Oliveira, “Biotechnology, Big Data and Artificial Intelligence,” *Biotechnol. J.*, vol. 14, no. 8, pp. 1–6, 2019.

[5] FDA, “PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance, Guidance for Industry,” 2004. [Online]. Available: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf.

[6] F. L. Vetter, S. Zobel-Roos, and J. Strube, “PAT for continuous chromatography integrated into continuous manufacturing of biologics towards autonomous operation,” *Processes*, vol. 9, no. 3, 2021.

[7] L. Rolinger, M. Rüdt, and J. Hubbuch, “A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing,” *Anal. Bioanal. Chem.*, vol. 412, no. 9, pp. 2047–2064, 2020.

[8] C. F. Mandenius and R. Gustavsson, “Mini-review: Soft sensors as means for PAT in the manufacture of bio-therapeutics,” *J. Chem. Technol. Biotechnol.*, vol. 90, no. 2, pp. 215–227, 2015.

[9] M. Degerman, K. Westerberg, and B. Nilsson, “Determining critical process parameters and process robustness in preparative chromatography - A model-based approach,” *Chem. Eng. Technol.*, vol. 32, no. 6, pp. 903–911, 2009.

[10] EMA, “Guideline on Real Time Release Testing (formerly Guideline on Parametric Release),” *European Medicines Agency*. pp. 1–10, 2012.

[11] M. Jiang *et al.*, “Opportunities and challenges of real-time release testing in biopharmaceutical manufacturing,” *Biotechnol. Bioeng.*, vol. 114, no. 11, pp. 2445–2456, 2017.

[12] S. Zobel-Roos *et al.*, “Accelerating biologics manufacturing by modeling or: Is Approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin?,” *Processes*, vol. 7, no. 2, pp. 1–28, 2019.

[13] A. Jungbauer and G. Carta, *Protein Chromatography: Process Development and Scale-Up*. Wiley-VCH Verlag GmbH & Co. KGaA, 2010.

[14] A. S. Rathore, R. Wood, A. Sharma, and S. Dermawan, “Case study and application of process analytical technology (PAT) towards bioprocessing: II. Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography,” *Biotechnol. Bioeng.*, vol. 101, no. 6, pp. 1366–1374, 2008.

[15] A. Armstrong *et al.*, “Advanced control strategies for bioprocess chromatography: Challenges and opportunities for intensified processes and next generation products,” *J. Chromatogr. A*, vol. 1639, p. 461914, 2021.

[16] L. K. Shekhawat and A. S. Rathore, “Mechanistic modeling based process analytical technology implementation for pooling in hydrophobic interaction chromatography,” *Biotechnol. Prog.*, vol. 35, no. 2, p. e2758, 2019.

[17] A. S. Rathore, M. Yu, S. Yeboah, and A. Sharma, “Case study and application of process analytical technology (PAT) towards bioprocessing: Use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography,” *Biotechnol. Bioeng.*, vol. 100, no. 2, pp. 306–316, 2008.

[18] D. G. Sauer *et al.*, “A two-step process for capture and purification of human basic fibroblast growth factor from E. coli homogenate: Yield versus endotoxin clearance,” *Protein Expr. Purif.*, vol. 153, 2019.

[19] D. G. Sauer *et al.*, “Real-time monitoring and model-based prediction of purity and quantity during a chromatographic capture of fibroblast growth factor 2,” *Biotechnol. Bioeng.*, vol. 116, no. 8, pp. 1999–2009, 2019.

[20] B. Sissolak, N. Lingg, W. Sommeregger, G. Striedner, and K. Vorauer-Uhl, “Impact of mammalian cell culture conditions on monoclonal antibody charge heterogeneity: an accessory monitoring tool for process development,” *J. Ind. Microbiol. Biotechnol.*, vol. 46,

no. 8, pp. 1167–1178, 2019.
[21] J. Coghlan, H. He, and A. S. Schwendeman, “Overview of Humira® Biosimilars: Current European Landscape and Future Implications,” *J. Pharm. Sci.*, vol. 110, no. 4, pp. 1572–1582, 2021.
[22] B. Hintersteiner, N. Lingg, E. Janzek, O. Mutschlechner, H. Loibner, and A. Jungbauer, “Microheterogeneity of therapeutic monoclonal antibodies is governed by changes in the surface charge of the protein,” *Biotechnol. J.*, vol. 11, no. 12, pp. 1617–1627, 2016.
[23] N. Walch *et al.*, “Prediction of the Quantity and Purity of an Antibody Capture Process in Real Time,” *Biotechnol. J.*, vol. 14, no. 7, pp. 1–10, 2019.
[24] A. Christler *et al.*, “Semi-automation of process analytics reduces operator effect,” *Bioprocess Biosyst. Eng.*, vol. 43, no. 5, pp. 753–764, 2020.
[25] D. S. Nascimento *et al.*, “MIQuant – Semi-Automation of Infarct Size Assessment in Models of Cardiac Ischemic Injury,” *PLoS One*, vol. 6, no. 9, p. e25045, 2011.
[26] S. Steiner *et al.*, “Organic synthesis in a modular robotic system driven by a chemical programming language,” *Science (80-.).*, vol. 363, no. 6423, 2019.
[27] J. L. J. Van De Bilt, J. H. J. Derks, and J. D. Janse, “Automated, high-throughput immunofluorescence staining: A new approach,” *Eur. J. Plant Pathol.*, vol. 120, no. 1, pp. 91–96, 2008.
[28] R. J. Francis, M. B. Lyons, R. T. Kingsford, and K. J. Brandis, “Counting mixed breeding aggregations of animal species using drones: Lessons from waterbirds on semi-automation,” *Remote Sens.*, vol. 12, no. 7, 2020.
[29] D. Wei, M. B. Hall, A. Sherehiy, and D. O. Popa, “Design and Evaluation of Human–Machine Interface for NEXUS: A Custom Microassembly System,” *J. Mico Nano-Manufacturing*, vol. 8, no. 4, 2020.
[30] A. Gates, M. Gates, M. Sebastianski, S. Guitard, S. A. Elliott, and L. Hartling, “The semi-automation of title and abstract screening: A retrospective exploration of ways to leverage Abstrackr’s relevance predictions in systematic and rapid reviews,” *BMC Med. Res. Methodol.*, vol. 20, no. 1, pp. 1–9, 2020.
[31] C. D. Hawker, “Nonanalytic laboratory automation: A quarter century of progress,” *Clin. Chem.*, vol. 63, no. 6, pp. 1074–1082, 2017.
[32] M. G. Björksten, B. Almby, and E. S. Jansson, “Hand and shoulder ailments among laboratory technicians using modern plunger-operated pipettes,” *Appl. Ergon.*, vol. 25, no. 2, pp. 88–94, 1994.
[33] H. Fleischer and K. Thurow, *Automation Solutions for Analytical Measurements: Concepts and Application*, First. Wiley-VCH Verlag GmbH & Co. KGaA, 2018.
[34] V. Rajamanickam, M. Krippl, C. Herwig, and O. Spadiut, “An automated data-driven DSP development approach for glycoproteins from yeast,” *Electrophoresis*, vol. 38, no. 22–23, pp. 2886–2891, 2017.
[35] A. H. Lou, M. O. Elnenaei, I. Sadek, S. Thompson, B. D. Crocker, and B. Nassar, “Evaluation of the impact of a total automation system in a large core laboratory on turnaround time,” *Clin. Biochem.*, vol. 49, no. 16–17, pp. 1254–1258, 2016.
[36] C. Czeslik, H. Seemann, and R. Winter, *Basiswissen Physikalische Chemie*. Vieweg+Teubner Verlag, 2009.
[37] P. Biechele, C. Busse, D. Solle, T. Scheper, and K. Reardon, “Sensor systems for bioprocess monitoring,” *Eng. Life Sci.*, vol. 15, no. 5, pp. 469–488, 2015.
[38] D. P. Wasalathanthri *et al.*, *Technology outlook for real-time quality attribute and process parameter monitoring in biopharmaceutical development—A review*, vol. 117, no. 10. 2020.
[39] J. Claßen, F. Aupert, K. F. Reardon, D. Solle, and T. Scheper, “Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application,” *Anal. Bioanal. Chem.*, vol. 409, no. 3, pp. 651–666, 2017.
[40] A. S. Rathore and G. Kapoor, “Application of process analytical technology for downstream purification of biotherapeutics,” *J. Chem. Technol. Biotechnol.*, vol. 90, no. 2, pp. 228–236, 2015.
[41] M. Rüdt, T. Briskot, and J. Hubbuch, “Advances in downstream processing of biologics – Spectroscopy: An emerging process analytical technology,” *J. Chromatogr. A*, vol. 1490, pp. 2–9, 2017.
[42] A. Christler *et al.*, “Technology transfer of a monitoring system to predict product concentration and purity of biopharmaceuticals in real-time during chromatographic separation,” *Biotechnol. Bioeng.*, 2021.
[43] B. A. Patel *et al.*, “Multi-angle light scattering as a process analytical technology measuring

real-time molecular weight for downstream process control," *MAbs*, vol. 10, no. 7, pp. 945–950, 2018.
[44] M. Gey, *Instrumentelle Analytik und Bioanalytik*, 3rd ed. Springer Berlin Heidelberg, 2015.
[45] J. Kastenhofer, J. Libiseller-Egger, V. Rajamanickam, and O. Spadiut, "Monitoring E. coli Cell Integrity by ATR-FTIR Spectroscopy and Chemometrics: Opportunities and Caveats," *Processes*, vol. 9, no. 422, 2021.
[46] P. Doppler, C. Gasser, R. Kriechbaum, A. Ferizi, and O. Spadiut, "In Situ Quantification of Polyhydroxybutyrate in Photobioreactor Cultivations of Synechocystis sp. Using an Ultrasound-Enhanced ATR-FTIR Spectroscopy Probe," *Bioengineering*, vol. 8, no. 9, p. 129, 2021.
[47] R. C. Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2020.
[48] F. A. Chiappini, C. M. Teglia, Á. G. Forno, and H. C. Goicoechea, "Modelling of bioprocess non-linear fluorescence data for at-line prediction of etanercept based on artificial neural networks optimized by response surface methodology," *Talanta*, vol. 210, no. November 2019, p. 120664, 2020.
[49] M. D. Peris-Díaz and A. Krężel, "A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry," *TrAC - Trends Anal. Chem.*, vol. 135, 2021.
[50] F. Feidl *et al.*, "A new flow cell and chemometric protocol for implementing in-line Raman spectroscopy in chromatography," *Biotechnol. Prog.*, vol. 35, no. 5, pp. 1–10, 2019.
[51] D. A. M. Pais, R. M. C. Portela, M. J. T. Carrondo, I. A. Isidro, and P. M. Alves, "Enabling PAT in insect cell bioprocesses: In situ monitoring of recombinant adeno-associated virus production by fluorescence spectroscopy," *Biotechnol. Bioeng.*, vol. 116, no. 11, pp. 2803–2814, 2019.
[52] P. Kroll, A. Hofer, S. Ulonska, J. Kager, and C. Herwig, "Model-Based Methods in the Biopharmaceutical Process Lifecycle," *Pharm. Res.*, vol. 34, no. 12, pp. 2596–2613, 2017.
[53] M. von Stosch *et al.*, "Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry," *Biotechnol. J.*, vol. 9, no. 6, pp. 719–726, 2014.
[54] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer New York, 2013.
[55] K. Pandya, C. A. Ray, L. Brunner, J. Wang, J. W. Lee, and B. DeSilva, "Strategies to minimize variability and bias associated with manual pipetting in ligand binding assays to assure data quality of protein therapeutic quantification," *J. Pharm. Biomed. Anal.*, vol. 53, no. 3, pp. 623–630, 2010.
[56] G. Lippi, G. Lima-Oliveira, G. Brocco, A. Bassi, and G. L. Salvagno, "Estimating the intra- and inter-individual imprecision of manual pipetting," *Clin. Chem. Lab. Med.*, vol. 55, no. 7, pp. 962–966, 2017.
[57] P. L. Tornel, E. Ayuso, and P. Martinez, "Evaluation of the turnaround time of an integrated preanalytical and analytical automated modular system in a medium-sized laboratory," *Clin. Biochem.*, vol. 38, no. 6, pp. 548–551, 2005.
[58] V. Patel *et al.*, "Automating bioanalytical sample analysis through enhanced system integration," *Bioanalysis*, vol. 5, no. 13, pp. 1649–1659, 2013.
[59] A. B. Ahene *et al.*, "Ligand binding assays in the 21st century laboratory: Automation," *AAPS J.*, vol. 14, no. 1, pp. 142–153, 2012.
[60] S. A. P. Pereira, P. J. Dyson, and M. L. M. F. S. Saraiva, "Miniaturized technologies for high-throughput drug screening enzymatic assays and diagnostics – A review," *TrAC - Trends Anal. Chem.*, vol. 126, p. 115862, 2020.
[61] Á. Ríos and M. Zougagh, "Modern qualitative analysis by miniaturized and microfluidic systems," *TrAC - Trends Anal. Chem.*, vol. 69, pp. 105–113, 2015.
[62] H. Salimi-Moosavi, P. Rathanaswami, S. Rajendran, M. Toupikov, and J. Hill, "Rapid affinity measurement of protein–protein interactions in a microfluidic platform," *Anal. Biochem.*, vol. 426, no. 2, pp. 134–141, 2012.
[63] S. Budroni *et al.*, "Computational modeling of microfluidic data provides high-throughput affinity estimates for monoclonal antibodies," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 3664–3672, 2021.
[64] J. Eid, M. Mougel, and M. Socol, "Advances in Continuous Microfluidics-Based Technologies for the Study of HIV Infection," *Viruses*, vol. 12, no. 9, pp. 1–16, 2020.
[65] I. Banerjee *et al.*, "Microfluidic centrifugation assisted precipitation based DNA quantification," *Lab Chip*, vol. 19, no. 9, pp. 1657–1664, 2019.

[66] P. Fernandes, F. Carvalho, and M. P. C. Marques, “Miniaturization in biotechnology: speeding up the development of bioprocesses.,” *Recent Pat. Biotechnol.*, vol. 5, no. 3, pp. 160–73, 2011.
[67] D. Totaro *et al.*, “Microscale Perfusion-Based Cultivation for Pichia pastoris Clone Screening Enables Accelerated and Optimized Recombinant Protein Production Processes,” *Biotechnol. J.*, vol. 16, no. 3, 2021.
[68] S. Großhans *et al.*, “In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography,” *J. Chromatogr. A*, vol. 1547, pp. 37–44, 2018.
[69] N. Brestrich, M. Rüdt, D. Büchler, and J. Hubbuch, “Selective protein quantification for preparative chromatography using variable pathlength UV/Vis spectroscopy and partial least squares regression,” *Chem. Eng. Sci.*, vol. 176, pp. 157–164, 2018.
[70] M. Pathak, K. Lintern, V. Chopda, D. G. Bracewell, and A. S. Rathore, “Fluorescence based real time monitoring of fouling in process chromatography,” *Sci. Rep.*, vol. 7, no. March, pp. 1–8, 2017.
[71] J. C. König *et al.*, “Development and characterisation of a new fluorescence sensor for online monitoring of bioprocesses,” *J. Sensors Sens. Syst.*, vol. 7, no. 2, pp. 461–467, 2018.
[72] H. Helgers *et al.*, “Towards autonomous operation by advanced process control—process analytical technology for continuous biologics antibody manufacturing,” *Processes*, vol. 9, no. 1, pp. 1–31, 2021.
[73] V. Kumar and A. S. Rathore, “Mechanistic Modeling Based PAT Implementation for Ion-Exchange Process Chromatography of Charge Variants of Monoclonal Antibody Products,” *Biotechnol. J.*, vol. 12, no. 9, pp. 1–9, 2017.
[74] W. K. Marek, D. Sauer, A. Dürauer, A. Jungbauer, W. Piątkowski, and D. Antos, “Prediction tool for loading, isocratic elution, gradient elution and scaling up of ion exchange chromatography of proteins,” *J. Chromatogr. A*, vol. 1566, pp. 89–101, 2018.
[75] M. von Stosch, R. Oliveira, J. Peres, and S. Feyo de Azevedo, “Hybrid semi-parametric modeling in process systems engineering: Past, present and future,” *Comput. Chem. Eng.*, vol. 60, pp. 86–101, 2014.

8. Contributions to publications

Publication I: Semi-automation of process analytics reduces operator effect

Bioprocess and Biosystems Engineering (2020) 43:753-764; Springer.

Anna Christler, Edit Felföldi, Magdalena Mosor, Dominik Sauer, Nicole Walch, Astrid Dürauer, Alois Jungbauer

https://doi.org/10.1007/s00449-019-02254-y

I, Anna Christler developed the concept of the transfer and comparison of manual analytical methods to semi-automated methods. I acquired parts of the data, I evaluated all data and wrote and edited the manuscript.

Publication II: Technology transfer of a monitoring system to predict product quantity and purity of biopharmaceuticals in real-time during chromatographic separation

Biotechnology & Bioengineering (2021) 1-12; Wiley.

Anna Christler*, Theresa Scharl-Hirsch*, Dominik Sauer, Johannes Köppl, Cabir Toy, Michael Melcher, Alois Jungbauer, Astrid Dürauer

*equally contributing authors

https://doi.org/10.1002/bit.27870

I, Anna Christler, set-up the real-time monitoring system at one other site together with Dominik Sauer and Johannes Köppl. I acquired parts of the data, evaluated the data together with the co-authors, wrote and edited the manuscript. Theresa Scharl, as equally contributing author, built and applied the prediction models and edited the manuscript.

Additional publication: A two-step process for capture and purification of human basic fibroblast growth factor from *E.coli* homogenate: Yield versus endotoxin clearance

Protein Expression and Purification (2019) 153, 70-82; Elsevier.

Dominik Georg Sauer, Magdalena Mosor, Anna-Carina Frank, Florian Weiß, **Anna Christler**, Nicole Walch, Alois Jungbauer, Astrid Dürauer

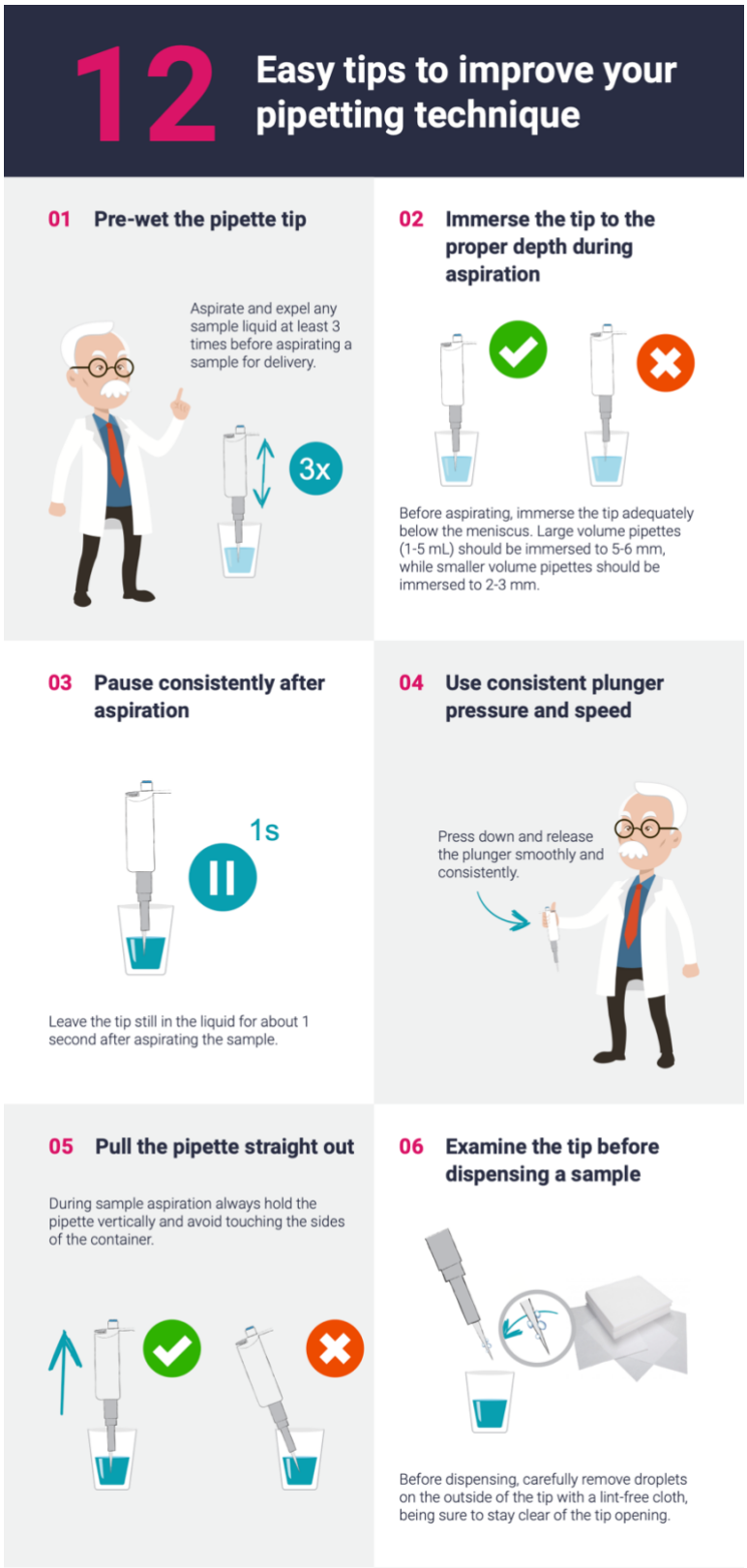
https://doi.org/10.1016/j.pep.2018.08.009

I, Anna Christler, established the endotoxin assay in the lab and acquired the endotoxin data.

9. Appendix

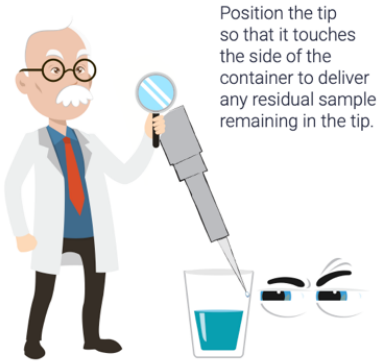
9.1. Infographic on Good Pipetting Technique

From https://biosistemika.com/blog/tips-to-improve-pipetting-technique/, accessed September 19, 2021.



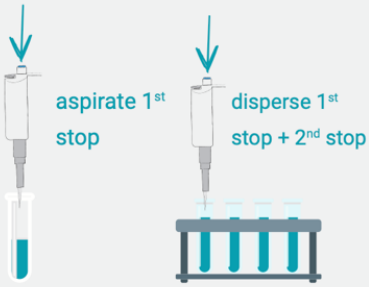
(continued on next page)

07 Examine the tip after dispensing a sample



Position the tip so that it touches the side of the container to deliver any residual sample remaining in the tip.

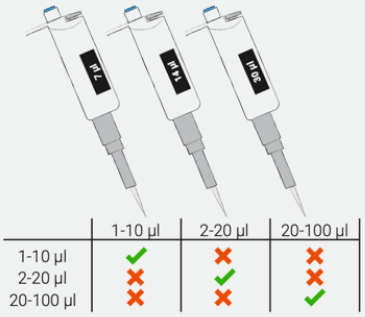
08 Use standard mode pipetting



Choose "standard (or forward) mode" pipetting rather than "reverse mode" for all aqueous samples.

09 Use the appropriate pipette

Pipette with a volume range closest to the volume you plan to aspirate and dispense.



	1-10 µl	2-20 µl	20-100 µl
1-10 µl	✓	✗	✗
2-20 µl	✗	✓	✗
20-100 µl	✗	✗	✓

10 Use the correct pipette tip

Use high quality tips intended for use with specific pipettes.

11 Work at ambient temperature equilibrium

Allow liquids and equipment to reach an equilibrium at an ambient temperature before you begin pipetting.

12 Minimize pipette handling

Hold the pipette loosely, return it to the pipette stand or set it down when you are not pipetting.

Wear gloves to reduce body heat transfer to the pipette.

Struggling with pipetting consistency?
Don't worry, use Pipetting Aid PlatR

BioSistemika

9.2. Publications

Bioprocess and Biosystems Engineering (2020) 43:753–764
https://doi.org/10.1007/s00449-019-02254-y

RESEARCH PAPER

Semi-automation of process analytics reduces operator effect

A. Christler[1] · E. Felföldi[1] · M. Mosor[1] · D. Sauer[1] · N. Walch[1] · A. Dürauer[1,2] · A. Jungbauer[1,2]

Received: 11 September 2019 / Accepted: 11 November 2019 / Published online: 7 December 2019
© The Author(s) 2019

Abstract
The aim of this study was to semi-automate process analytics for the quantification of common impurities in downstream processing such as host cell DNA, host cell proteins and endotoxins using a commercial liquid handling station. By semi-automation, the work load to fully analyze the elution peak of a purification run was reduced by at least 2.41 h. The relative standard deviation of results among different operators over a time span of up to 6 months was at the best reduced by half, e.g. from 13.7 to 7.1% in dsDNA analysis. Automation did not improve the reproducibility of results produced by one operator but released time for data evaluation and interpretation or planning of experiments. Overall, semi-automation of process analytics reduced operator-specific influence on test results. Such robust and reproducible analytics is fundamental to establish process analytical technology and get downstream processing ready for Quality by Design approaches.

Keywords Liquid handling · Pipetting · PicoGreen · Endotoxin · Host cell proteins · dsDNA · ELISA

Introduction

Semi-automation is the compromise to accelerate process development while maintaining high precision and reasonable costs in an environment where the number of samples is manageable. The benefits of full automation are often overestimated. Full automation is desirable if highest productivity is anticipated and to eliminate most human influence. Since the beginning of the 1990s, especially after the turn of the millennium, automated liquid handling systems enabled high-throughput methods and thus revolutionized laboratory work. This is highlighted by numbers of publications in different fields of application such as nucleic acid synthesis and analysis, protein refolding, production host clone screening, process development, diagnostics, cell culture and others [1-9]. However, setting up fully automated methods is laborious, time-consuming [10, 11] and investment costs for fully automated equipment are high. This implies a long-term commitment for a specific assay. Usually 3–5 years is necessary for full automation to depreciate the high upfront costs for the equipment. Chan emphasizes that high-volume testing, meaning many samples, may benefit using semi-automated steps for the most labor-intensive steps [12].

Conventional as well as DoE-based process development, process modelling, and manufacturing processes of recombinant proteins require in-process analytics to monitor the quality of the product and depletion of impurities. Especially, biopharmaceuticals are subjected to stringent regulatory requirements. Appropriate analytical methods must provide information about the content of target protein, its activity and show that critical impurities such as DNA, host cell proteins (HCP), endotoxins and product-related impurities are cleared from the final protein product below certain levels as specified by national and international authorities. Accuracy and precision of analytical results are usually assured by method qualification, validation and continuous training of operators. However, Pandya et al. found that their long-term stability testing of protein therapeutics was obscured by the systematic differences in manual pipetting between operators [13], commonly known as "operator effect". Moreover, manual pipetting is highly repetitive and might lead to fatigue and possibly repetitive strain injury [14]. Therefore, we hypothesize that an appropriate

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00449-019-02254-y) contains supplementary material, which is available to authorized users.

✉ A. Jungbauer
alois.jungbauer@boku.ac.at

[1] Austrian Centre of Industrial Biotechnology, Muthgasse 11, 1190 Vienna, Austria

[2] Institute of Bioprocess Science and Engineering, Department of Biotechnology, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria

automation strategy will improve reproducibility of results in a long-term study involving multiple operators, increase security of data delivery due to an operator-independent analytical workflow and protect staff from adverse effects of monotonous work. Manpower released by automating repetitive work can be deployed for more demanding tasks and thus increase productivity.

In the present literature, either very specialized (semi-automated) applications were reported or liquid handling stations (LHS) with high capacities and functionalities were used. We extend the field by describing semi-automated methods for four different common biochemical assays (dsDNA, HCP, endotoxins, binding affinity) using a simple commercial LHS and comparing them to manual assay performance. The adaptation of the analytical protocols to the space and tools of the LHS is described. Spectroscopically detectable model substances were used to discriminate the influence of sample dilution steps only to deviations obtained for the individual assays. The outcome of the semi-automated protocols was compared to conventional manual analytics in terms of time efficiency, precision and reproducibility. Semi-automation of analytics reduced long-term variability of analytical results and improved the confidence in the subsequent data application.

Materials and methods

Instrumentation

A liquid handling station epMotion®5073 (Eppendorf, Germany) equipped with a thermal module (0–110 °C), a single-channel liquid transfer tool (40–1000 µL) and an 8-channel tool (20–300 µL) was used. Plates were incubated in a Thermomixer Comfort MTP (Eppendorf, Germany). Spectroscopic measurements were performed using a plate reader Infinite M200 PRO (TECAN, USA) with a dual pump dispense unit. Preparative chromatography was carried out on an ÄKTA pure 25 workstation (GE Healthcare, USA).

Chemicals

Chemicals were purchased from E. Merck (Germany) in analytical grade unless specified differently. Green fluorescent protein (GFP) was produced in-house and kindly provided by Prof. Rainer Hahn.

Protein samples

An IgG1 monoclonal antibody (mAb) was produced in CHO cell culture, harvested, and captured by Protein A affinity chromatography as described in [15]. IgG1 concentrations were determined by high-performance monolith affinity chromatography as described in [16]. Human fibroblast growth factor 2 (FGF-2) was expressed in *E. coli*, captured by cation exchange or affinity chromatography, polished by hydrophobic interaction chromatography, and quantified by reversed phase HPLC, all as described in [17].

Biochemical assays

The principles described below are valid for both manual and semi-automated procedures. Serial 1:2 dilutions were done in 350 µL NUNC® 96F 96-well microplates (Thermo Fisher Scientific, USA). From the dilution plates, 100 µL were transferred to the measurement plates (Corning® Costar 350 µl 96-well plates, Sigma-Aldrich/Merck, USA) or MaxiSorp™ Immuno ELISA plates (Thermo Fisher Scientific, USA). Data were evaluated using MS Excel (Microsoft, USA). In semi-automated methods, samples for binding affinity and endotoxins were pre-diluted in 96-well 2 mL deepwell® plates (VWR, USA). In manual methods, the dilution was carried out in 1.5 mL (Sarstedt, Germany) and 2 mL reaction tubes (Eppendorf, Germany).

Host cell protein (HCP) ELISA

Capture and detection antibodies and HCP standards were purchased from Cygnus, USA. Product numbers are given in brackets. ELISA plates were coated with 0.25 µg of anti-*E. coli* HCP (AP117) or 0.5 µg anti-CHO HCP (3G-0016-AF) antibody per well in 100 µL of 0.2 M sodium carbonate buffer (pH 9.3–9.5) for 2 h at 37 °C/350 rpm. Plates were washed three times with 300 µL of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) containing 0.05% Tween 20 (pH 7.2–7.6) per well. Plates were blocked with 300 µL 3% BSA in PBS per well overnight at 4 °C. The blocked plates were washed as before. Samples and concentrated *E. coli* or CHO HCP antigen (F413H or F553H) were diluted in sample buffer (1% BSA, 0.05% Tween 20 in PBS) and incubated for 1 h at 37 °C/350 rpm. Plates were washed as before and incubated with 100 µL/well of a 0.5 µg/mL (0.05 µg/well) detection antibody solution (anti-*E. coli*-HCP, F411C or anti-CHO–HCP, F551C) conjugated with horseradish peroxidase (HRP) in sample buffer for 1 h at 37 °C/350 rpm. Plates were washed again as before and incubated with 100 µL/well of a tetramethylbenzidine (TMB) substrate (Bio-Rad, USA) for 30 min at room temperature without shaking. The HRP reaction was stopped by adding 50 µL/well of 1 N sulfuric acid. Absorbance was measured at 450 nm and at 630 nm as reference which was subtracted from the absorbance at 450 nm. Average blank was subtracted from all measurements. A quadratic calibration curve was fitted through the standard measurements. The calibration range for *E. coli* HCP was 0.39–25 ng/mL and 2.11–135 ng/mL for CHO HCP.

Double-stranded (ds) DNA quantification by Quant-iT™ PicoGreen® assay

DsDNA concentrations were determined with Quant-iT™ PicoGreen® assay (Invitrogen, USA). 20 × TE buffer was diluted 1:20 with RO-water to a working concentration of 10 mM Tris–HCl, 1 mM EDTA, pH 7.5 (1 × TE). Samples and λ DNA standard were diluted in 1 × TE. 100 µL of Quant-iT™ PicoGreen® working solution in 1 × TE was added to each well. After incubation for 2 min at room temperature in the dark, fluorescence was measured using an excitation wavelength of 480 nm and emission wavelength of 520 nm (filter with a bandwidth of ± 20 nm). Average blank was subtracted from all measurements. A linear calibration curve was fitted through the standard measurements and the origin of the coordinate system (0,0). The calibration range for *E. coli* DNA was 3.91–500 ng/mL and 1.95–250 ng/mL for CHO DNA.

Endotoxin quantification with recombinant factor C-based assay

Endotoxins were determined using EndoZyme® II recombinant Factor C (rFC)-based assay kit (Hyglos, Germany). Samples and standards were diluted in endotoxin-free water. Vigorous mixing (30–120 s on orbital shaker at 1400 rpm or ten cycles of aspiration and dispense at a speed of 11 mm/s) was applied to disperse the analytes homogeneously. The plate was heated to 37 °C. 100 µL of enzyme–substrate solution was added to each sample and standard dilution. Signal intensities were measured at an excitation wavelength of 380 nm and emission wavelength of 445 nm. Plates were incubated at 37 °C for 75 min without shaking. Signals at time 0 were subtracted from signals after 75 min. Average blank was subtracted from all measurements. A linear calibration curve was fitted through the standard measurements and the origin of the coordinate system (0,0). The calibration range was 0.01–5 Endotoxin Units (EU)/mL.

Determination of ligand binding affinity with a surface plasmon resonance (SPR)-based assay

Binding affinities of anti-TNFα-IgG against TNFα (10,602-HNAE-100, Sino Biological, China) and of FGF-2 to FGF-receptor 2 were determined by a SPR assay using a Biacore 2000 system (GE Healthcare, USA) as described in [17].

Quality criteria

For the standard curve fit of PicoGreen®, HCP ELISA and endotoxin assays, a value of the determination coefficient R^2 of at least 0.999 was accepted. A maximum tolerable deviation from the nominal concentration (bias) of ± 15% was allowed. For each reported target response, at least three consecutive values from different dilutions were averaged that give a coefficient of variation (CV) ≤ 20%. The lower limit of detection (LLOD) was calculated as the average of at least three blank measurements plus three times the standard deviation of blanks. The lower limit of quantification (LLOQ) was calculated as average blank plus ten times the standard deviation of blanks. The upper limit of quantification (ULOQ) was the highest calibration standard.

Comparison of accuracy and precision of automated dilutions compared to manual

To compare the processes of sample and standard dilution only, model substances were used that can be detected spectroscopically. Solutions of myoglobin, the pH indicator bromocresol purple, and green fluorescent protein (GFP) were prepared in the respective buffer system. Dilutions of model substance solutions were treated like process samples to compare the manual and semi-automated methods as close to real situations as possible. Random concentration levels were assumed for the model solutions and dilutions calculated accordingly. Averages of 2 replicates for each dilution level were compared and the respective differences between manual and automated results plotted over the average absorbance value according to Bland and Altman, 1999 [18].

Results

Transfer of analytical methods from manual to semi-automated processes

Four analytical methods commonly applied for process analytics in downstream processing of recombinant proteins were semi-automated using a commercial LHS. The analytical methods comprised the quantification of host cell dsDNA, HCP, bacterial endotoxins and binding assays for potency estimation. The methods were adjusted for processes capturing an IgG1 antibody and a basic fibroblast growth factor-2 (FGF-2), respectively. Different steps of the analytical procedures were automated (Fig. 1): sample dilution, filling of diluted samples in vials (binding affinity), and addition of reagents to samples (dsDNA and HCP). Sample dilution was identified as highly potential for automation, since this step consumed the most operator time and was very repetitive. For example, in the early-stage purification, endotoxin concentrations in samples exceeded the ULOQ up to 37,600-fold (Tables 1 and 2). Even samples with analyte concentrations within the quantification ranges usually must be diluted to eliminate or reduce matrix influence which otherwise can impair accurate quantification.

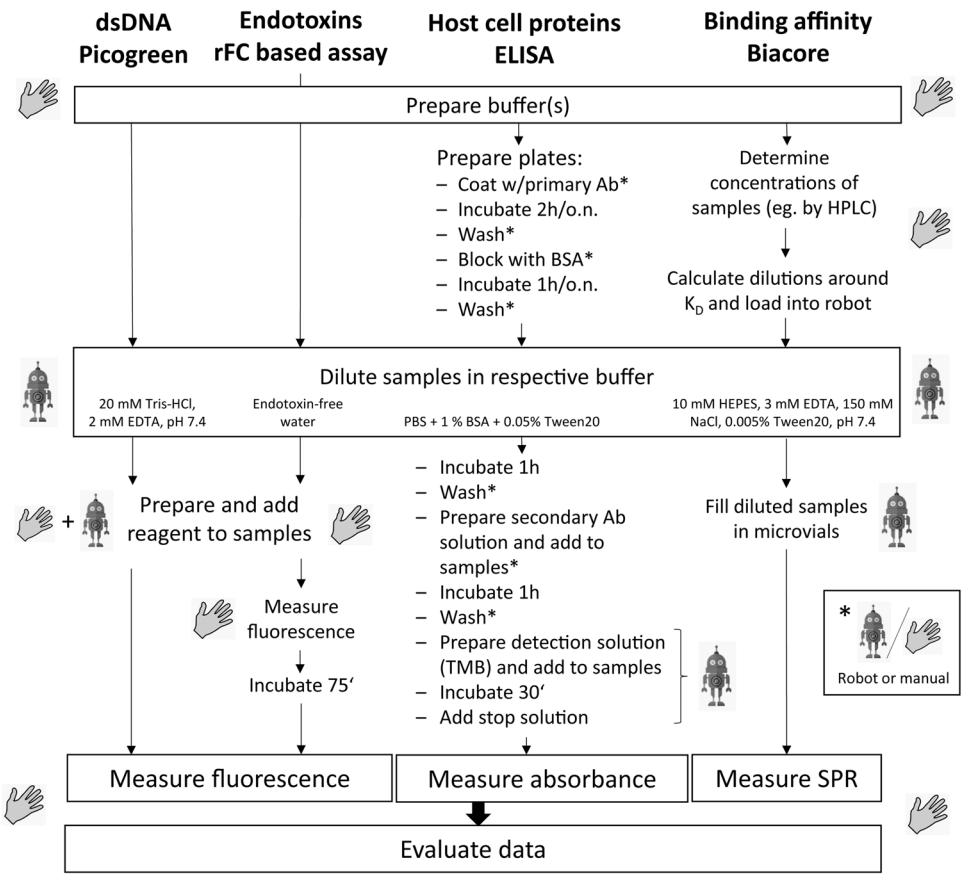


Fig. 1 Steps in the four semi-automated analytical procedures. Hands symbolize manual actions, robots denote automated steps. Asterisks denote steps with an automated option. Abbreviations: *rFC* recombinant factor C, *HCP* host cell proteins, *TMB* tetramethylbenzidin, *Ab* antibody, *HPLC* high-pressure liquid chromatography, *SPR* surface plasmon resonance

Table 1 Ranges of analytes in process samples from capture stage purification, analytical ranges and analyte concentrations relative to upper limits of quantification (ULOQ)

Protein sample	Component	Concentration range in fractions (mL^{-1})	Analytical range of assay (mL^{-1})	Analyte concentrations in % of ULOQ
FGF-2	Product	0.01–42 mg		
	DNA	50–2000 ng	3.91–500 ng	10–400%
	HCP	20–500 ng	0.39–25 ng	80–2000%
	Endotoxins	20–188,000 EU	0.01–5 EU	400–3760 000%
IgG	Product	0.004–32 mg		
	DNA	≤1.95*–100 ng	1.95–250 ng	≤40%
	HCP	0.060–30 µg	2.11–135 ng	44–22,222%

*Below lower limit of quantification (LLOQ)

Table 2 Ranges of analytes in process samples after FGF-2 polishing, analytical range of endotoxin assay and endotoxin concentrations relative to upper limit of quantification (ULOQ)

Protein sample	Component	Concentration range in fractions (mL^{-1})	Analytical range of assay (mL^{-1})	Analyte concentrations in % of ULOQ
FGF-2	Product	0.07–7.3 mg		
	DNA	<LLOQ		
	HCP	<LLOQ		
	Endotoxins	20–7000 EU	0.01–5 EU	400–140,000%

This was the case for DNA determination in samples of IgG1 where the measured concentrations never exceeded the ULOQ.

Differences between manual and semi-automated pipetting mainly concerned (1) the available space in the LHS, (2) the number and volume ranges of the pipetting tools and (3) the operating direction of the multichannel tool. In most of the methods, more units of labware (plates, reservoir holder, tube racks, tips) were needed compared to the available space on worktable of the LHS (Fig. S1 in Supplementary Material). In these cases, labware had to be interchanged manually during a method. The number of exchanges could be reduced using deepwell plates for sample dilution instead of reaction tubes. This allowed to divide dilution factors into smaller steps or to process more samples at a time.

Host cell dsDNA

DNA staining by Quant-iT™ PicoGreen® is a fast and sensitive method for dsDNA quantification. Automation of this assay was straightforward due to its short protocol and convenient pipetting behavior of the dilution buffer (low viscosity, high surface tension). Aliquot dispense of DNA staining reagent was semi-automated using a dispense unit attached to the spectroscopic plate reader to reduce light-induced degradation of the dye and exposure time of the operators to the hazardous reagent. The minimum dilution for all samples was 1:2.

Protein-specific binding affinity

To determine the binding affinity by SPR as a measure for the product´s potency, the protein concentration must be reduced to levels around the expected dissociation constant (K_D). Since the process samples contained different initial product concentrations (Tables 1 and 2), they must be normalized to the same concentration in the beginning of sample dilution. In the semi-automated procedure, this was achieved by manually generating csv-files based on templates and importing the volumetric information contained in them into the LHS, similar to the way described in [19]. Samples were filled in the dilution plate manually and all further dilution steps were executed automatically. Templates were created to calculate volumes to be used by the LHS (see Fig. S2A in Supplementary Material). This allowed fast and flexible sample processing in routine analytics.

To avoid disturbance of SPR measurements by gas bubbles, the running buffer was degassed by ultrasonication prior to sample dilution. Dissolution of air during sample dilution and filling as well as trapping of air bubbles in sample vials must be avoided. This was achieved by adjusting the speed of dispense to 3–4 mm/s. No higher frequency of gas bubble disturbance was observed in the sensorgrams of samples diluted automatically compared to manual sample dilution (data not shown). Therefore, we conclude that the robotic mixing by repeated aspiration and dispense did not dissolve more air than manual mixing on an orbital shaker.

First experiments indicated that measurement of SPR response and fitting of the data to a binding model caused more variance in the final analytical result than the sample dilution process. Therefore, model substances were used to assess the differences between manual and semi-automated sample dilution for binding affinity measurements. Buffered solutions of bromocresol purple and myoglobin were diluted in accordance to both protocols, manual and semi-automated. After normalization to a common concentration representing 100 nM antibody concentration, five consecutive independent dilutions with dilution factors between 1.3 and 10 were prepared. Limits of agreement defined as the 95% confidence intervals [18] were −5.8% and +7.0% for bromocresol purple (Fig. 2a, b) and −6.7% and +3% for myoglobin (Fig. 2c, d). The differences of the bromocresol purple sample contents were evenly distributed around zero. For the myoglobin solution, differences for smaller concentrations (higher dilution factors) were mostly negative and positive for the higher concentrations. Average differences of semi-automated results compared to manual dilution were +1.2% and −1.8% for bromocresol purple and myoglobin, respectively. These are acceptable ranges for use in the assay.

HCP ELISA

Immunological methods such as western blot and ELISA are the standard for HCP detection and quantification. HCP ELISA was the longest and most challenging protocol to semi-automate. One reason was that the buffers used contained BSA and Tween 20 to reduce unspecific binding of matrix components. Detergents facilitate air bubble formation in pipette tips upon aspiration and foaming on liquid surfaces upon dispense through a reduced surface tension and increased viscosity of the solution. Air bubbles and foam can lead to volume inaccuracies and must be avoided. On the other hand, pipetting speed must be high enough to ensure complete mixing. In our case, this was achieved by repeated (3–7x) aspiration and dispense at a speed of 1 mm/s. Exact values might vary with equipment.

A plate wash procedure involving four cycles of aspiration and dispense of wash solution was developed on the LHS. Equivalence of wash efficiency to manual wash was shown using GFP solution and measuring fluorescence after each wash cycle (Fig. 3). Three slightly different automated protocols were compared to manual: AUT1 and AUT2 include manual emptying of residual liquid in the plate after the wash procedure of three or four wash cycles, respectively.

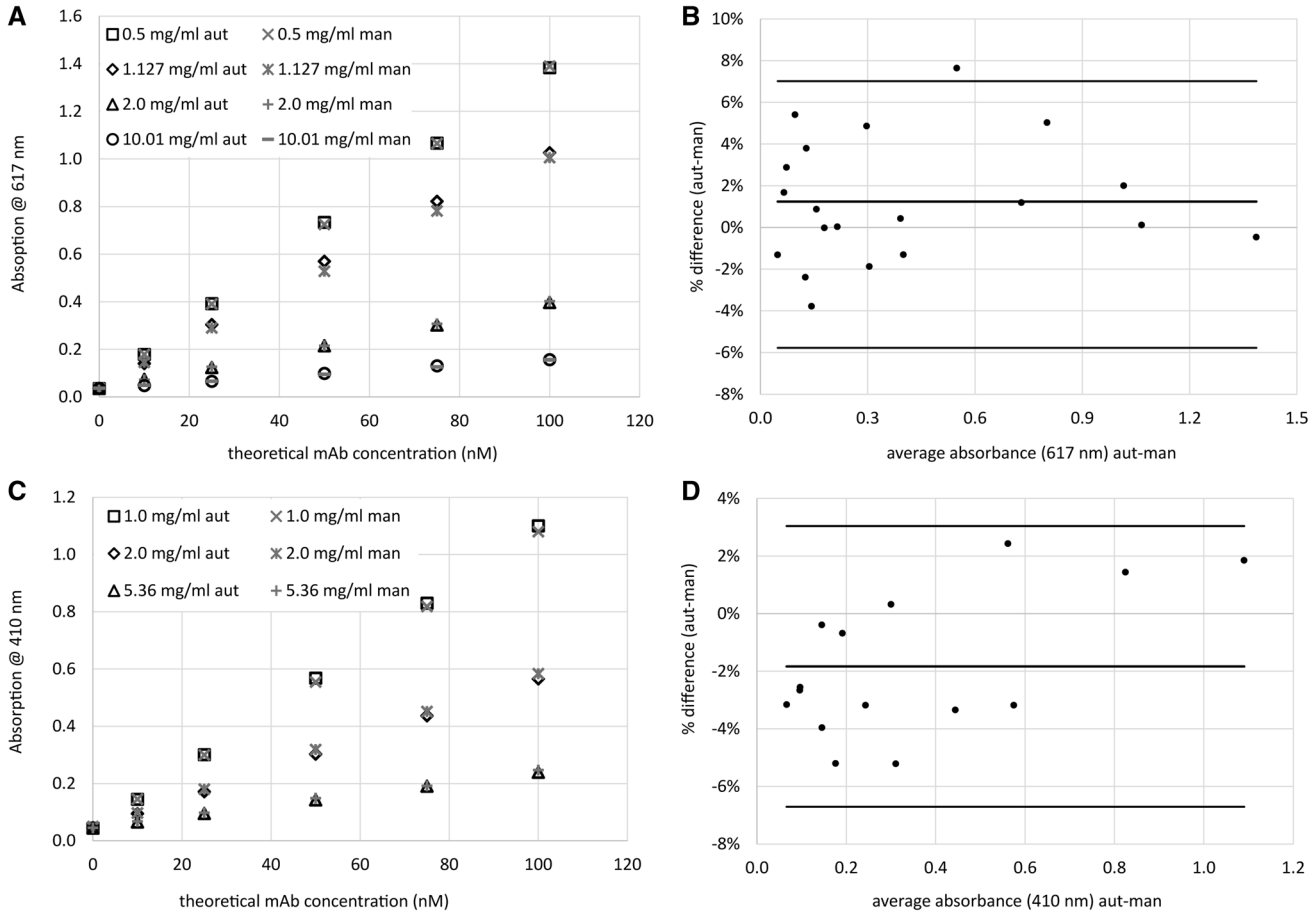


Fig. 2 Absorbance measurements of model substances bromocresol purple (**a**–**b**) and myoglobin (**c**–**d**) assuming different initial mAb concentrations. **b** and **d** show relative differences over the concentration ranges of the prepared samples with averages and limits of agreement

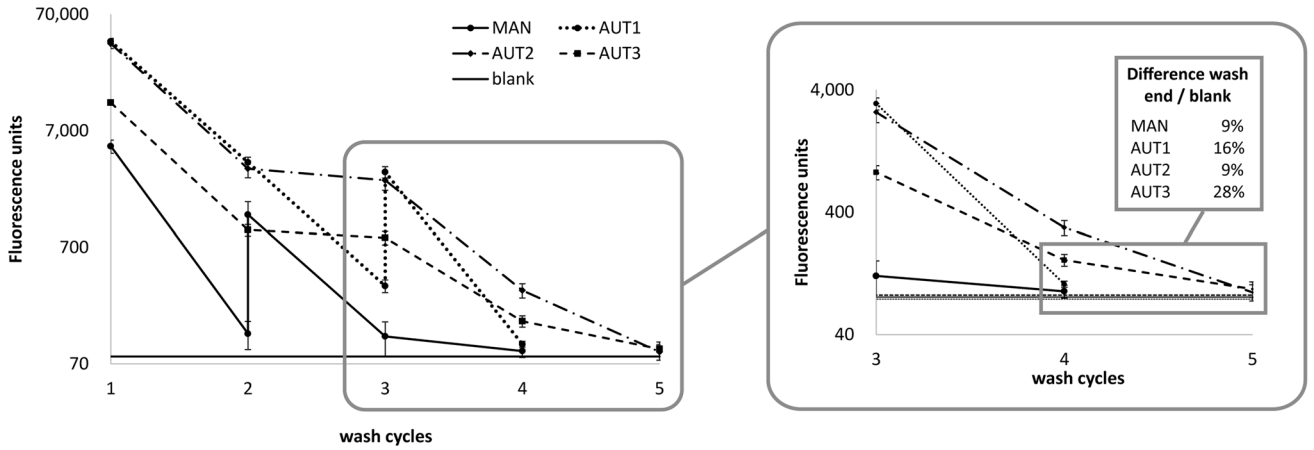


Fig. 3 Wash-out of GFP by manual and different semi-automated procedures monitored by fluorescence detection. Left panel: absolute fluorescence counts after each wash cycle. Signal amplification factor was increased when the signal dropped below 700 FU to distinguish signal from blank. Right panel: final fluorescence signals compared to blank (solid line ± one standard deviation)

For the third protocol (AUT3), manual emptying of residual liquid was done prior to four wash cycles. The residual fluorescence determined after the entire protocol was used to measure the efficiency of the procedures. Using AUT1 and AUT2, a depletion in the range of the manual protocol was achieved. AUT3 showed a 20% higher residual fluorescence than the manual protocol. AUT2 plate wash procedure was applied for HCP ELISA assays and compared to the manual protocol. Both plate wash methods produced equivalent results (data not shown), thus the semi-automated plate wash procedure with four wash cycles with manual emptying of residual liquid in the end was chosen as standard protocol.

Endotoxin

Efficient mixing is critical during sample preparation for endotoxin detection due to the amphiphilic nature of the lipopolysaccharides. According to the kit manufacturer, samples and standard solutions should be shaken for 1–2 min at 1400 rpm on an orbital shaker. Aspirating and dispensing solutions in the LHS 10 × at a speed of 9–14 mm/s resulted in equivalent signals of standard dilutions compared to manual mixing and dilution (data not shown). The manufacturer moreover recommended using glass containers rather than polymeric due to a potentially higher surface adherence of lipopolysaccharides to polymeric material [20, 21]. Recovery of endotoxins from glass and polymeric containers was tested by comparing signals of dilutions performed in the respective containers. Equivalent signals were obtained even at low concentrations (0.01–0.1 EU/mL). Full analyte recovery from polymeric containers enabled the dilution of samples and standard in polystyrene deepwell plates instead of in glass vials. Since in the early stage purification of *E.coli* homogenates, endotoxin levels were very high (up to 188 000 EU/ml), the number of vials required for dilution would have exceeded the available space in the LHS. Thus, the applicability of multiwell plates was an important prerequisite to semi-automate this assay.

Labware compatibility between manual and semi-automated methods

Most of the standard labware used in manual methods can be also used in the LHS as they are stored in a built-in database. Special pipette tips, reservoirs, reservoir holder and a tube rack were purchased with the instrument. Accurate information about geometries of the used labware were required for exact liquid aspiration and dispense. Starting liquid levels of buffers and samples were detected by the optical sensor of the LHS. The rise and fall of filling levels upon aspiration and dispense was then calculated by the system using the container geometry. Thus, a strongly deviating liquid level, due to for example deviating container geometry, can lead to distorted aspiration or dispense and thus produce wrong results. Therefore, specification of the labware is crucial for correct pipetting. Special labware can be sent to the LHS manufacturer to establish a dataset and use it in the LHS.

Pipetting tools

Plunger-operated pipettes usually achieve greatest accuracy and precision at the upper limit of their volume ranges. The accuracy of the single-channel tool was checked gravimetrically and spectroscopically. At the minimum volume recommended by the manufacturer for the single channel tool (40 µL), with ELISA sample buffer which is the fluid with the highest viscosity and the lowest surface tension of the solutions tested and considered similar to complex biological samples from early-stage purification (compare Sect. 3.1.1), the relative error was between + 3.3 and + 4.1% ($n = 6$). A minimal working volume for sample aspiration with the single-channel tool was set from experience at a value of 75 µL. For accurate serial dilutions, the 8-channel tool's precision was tested. At the upper limit (300 µL), an error of + 1.1 ± 0.8% ($n = 4$) was observed for pipetting of water and reuse of tips. This bias was reduced to − 0.4 ± 0.3% ($n = 4$) when fresh tips were used after each cycle which is in the specification range given by the vendor.

In the manual protocols, 12 specimens (samples, standard, blank, references) were diluted with a 12-channel pipette in vertical direction on dilution plates (Fig. 4a). In automated liquid handling systems, multichannel tools usually have eight channels and they are operated in horizontal direction. Various positionings of samples on dilution plates were applied to meet the needs of the different assays (Fig. 4b–e). Two parameters determined the most suitable type of arrangement: (1) the range of analytes in samples and (2) the analytical range and linearity of the assays. Assays with a large linear range, such as the PicoGreen® DNA assay and the recombinant Factor C-based endotoxin assay, required a lower number of measurements (e.g. four) to achieve results with sufficient accuracy. Nevertheless, due to the large range of endotoxin concentrations over the elution peak, five serial dilution levels per sample were used. After another purification step (FGF-2 polishing), less dilution levels were needed due to a more narrow range of concentrations in samples. The range of DNA levels in all samples was small enough so that four 1:2 serial dilutions per sample were sufficient. If the assay response increases non-linearly with analyte concentration or only in a small range as in the case of HCP ELISA, more dilutions are preferable. Due to the very high HCP concentrations compared to the analytical range and the assay non-linearity, six dilutions per sample were used. Schemes

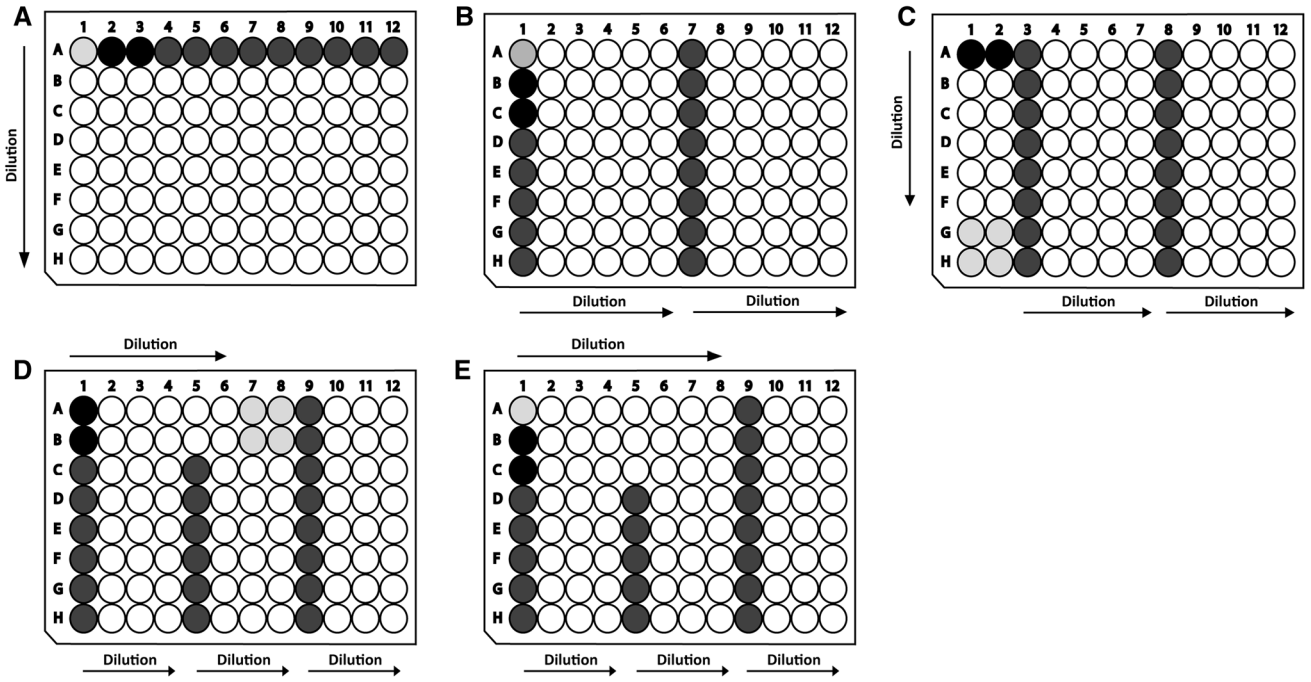


Fig. 4 Arrangements of samples on 96-well dilution plates used for different assays. Blank, reference standards and samples are placed in wells indicated in light, dark and medium grey, respectively, and diluted serially 1:2 in the direction of the arrows. **a** Application scheme in all manual assays. Automated assays: **b** ELISA, **c** endotoxins (capture), **d** endotoxins (polishing), **e** dsDNA

in Fig. 4c–e allowed to measure all fractions of a run on one plate. For HCP ELISA, 1.5 plates were required to analyze all fractions of one chromatographic run.

Performance comparison of semi-automated and manual procedures

Time

In a preparative chromatographic purification run in lab scale, typically 15–20 fractions are collected during elution which are then analyzed. In our assays for DNA, HCP and endotoxins, all samples were analyzed together, while for binding affinity determination, two times eight samples were determined consecutively, because many dilution steps were required. In Fig. 5, the total times of manual and semi-automated methods are compared for a polishing run. Details are given in the Supplementary Material, Table S1. The total times per assay were almost equal for manual and semi-automated methods with the exception of ELISA, where the semi-automated procedure requires relatively much operator time. Additionally, the speed of pipetting must be low in this assay due to the buffers' tendency to form air bubbles during pipetting.

The working time reduction by semi-automation of each assay was calculated as the difference between the sum of the manual steps in semi-automated method and the total time needed for manual execution. With the semi-automated procedures, operator working time could be reduced on average by 11.6–53.2 min per analysis. In total, for the analysis of product purity and potency, about 225 min or 3.75 h of sample dilution could be eliminated by semi-automation and around 144 min or 2.41 h of working time saved. The time saving will be even higher for earlier purification steps (product capture) where impurity levels are higher.

Accuracy and reproducibility

The inter-assay variations of DNA and HCP measurements were estimated using a quality control (QC) sample in each assay. Comparison of results from manual and semi-automated methods was sensible only for fully quantitative methods. Due to low DNA concentrations close to the LLOQ in IgG1 samples, matrix effects were dominant resulting in poor dilutional linearity. Poor dilutional linearity was also observed in HCP measurements of FGF-2 samples. We, therefore, considered these results as semi-quantitative and did not use them for accuracy and precision assessments. Measurements of the QC samples were normalized by division by the median of the respective manual assay results which were 76.29 µg/ml for DNA and 1.195 µg/ml for CHO HCP. The median of the semi-automated DNA measurements was 17% higher than the manual assay results while the distribution was narrower (Fig. 6a). For CHO HCP

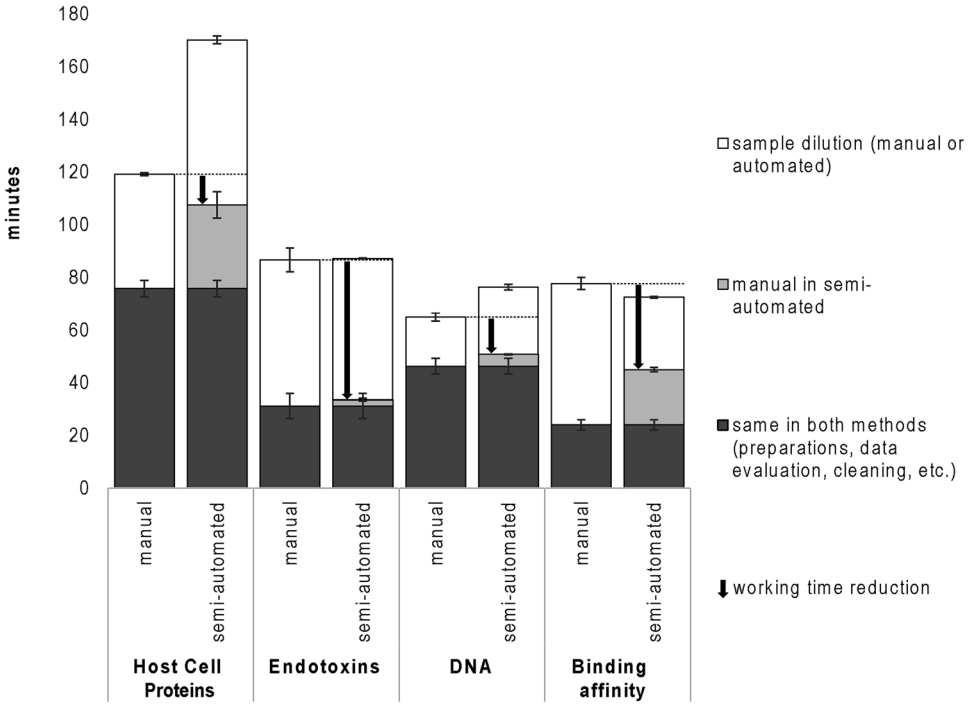


Fig. 5 Total and operator working times for typical sets of samples (all 15–20 fractions of one purification run for HCP, DNA and endotoxins, eight fractions for binding affinity). Error bars represent ± 1 standard deviation of three or four measurements. Dark grey parts are performed manually

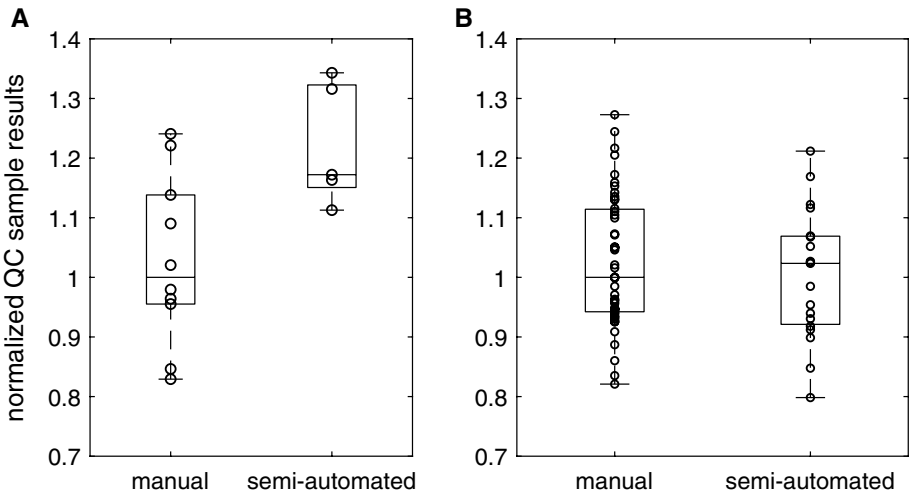


Fig. 6 Reproducibility in (**a**) PicoGreen DNA and (**b**) ELISA HCP assays

ELISA, a slightly larger inter-assay variability (10.9%) was found compared to manual (8.3%). The difference of average results was − 2% (Fig. 6b). Only one operator was involved producing the data, whereas for the DNA assay four operators were involved.

Discussion

Quality testing of products is a bottleneck in biopharmaceutical process development and manufacturing with its stringent regulatory requirements. Specifically, the laborious and time-consuming repetitive sample dilution are driving forces for the implementation of automated equipment. In this work, biochemical analytical protocols for the analysis of samples from chromatographic protein purification steps were semi-automated to reduce operator hands-on-time, parallelize workflows and accelerate development projects which depend on analytical results. Semi-automated protocols were developed starting from manual procedures and adjusted for typical protein samples from different expression hosts and purification processes, containing large ranges of analyte concentrations. Equivalent results were obtained compared to established manual methods except for DNA measurements, where higher values were obtained. The use of model substances allowed to

directly compare the dilution processes and exclude variance added by the detection method. This approach was most useful in methods using measurements of SPR or enzyme activity because these steps caused more variance to the result than the sample dilution process. Enzyme activity depends on temperature which is not perfectly distributed in 96-well plates and in thermoshakers [22] and might influence the final results.

Operator working times were reduced to very different extents for the different assays by the semi-automated procedures. The largest benefits in this respect were observed for binding affinity and endotoxin determinations, because large dilution factors and rigorous mixing was required.

ELISA plate wash was mostly done manually, because the automated plate wash took around 6.7 times as long as manual plate wash (around 40 min for 2 plates compared to around 6 min manually). Since the automated and the manual protocols led to results of same accuracy and reproducibility, plate wash was assigned optionally automated or manual in the semi-automated procedure depending on the individual time schedule and preference of the operator. Similarly, reagent dispense was left to the operators' discretion to carry out manually or semi-automated, since the methods were equal in terms of time and quality. Automation of reagent dispense with the liquid handling station was, however, beneficial in routine analytics when several steps could be combined such as staining, incubation and stop in ELISA (Fig. 1). Also, dispensing light-sensitive reagents such as PicoGreen® automatically in the plate reader reduces light-induced degradation and thus increases sensitivity compared to manual dispense. The reduced handling with the potentially mutagenic reagent moreover improved working safety.

Semi-automation lead to a reduction of operator hands-on-time and influence, but it required adjustments to overcome technical limitations of the LHS such as restricted space and pipetting tools. Generally, larger volumes were used compared to manual to increase precision. Compatibility of the LHS with common laboratory containers and plates is advantageous since it reduces dependency on any special materials and products from a designated vendor. Deepwell plates allowed to perform more sample dilutions on the same footprint compared to using reaction tubes. Thereby, the number of manual exchanges during a method could be reduced. More efficient sample patterns on multiwell plates could be carried out with the LHS. Thereby, the required quantities of reagents and materials as well as effort for data evaluation were reduced. Extensive repetitive movements as required for operating plunger-driven manual pipettes and thus the risk of repetitive strain injuries in the hands, arms and shoulders [14, 23] were reduced by automation.

The design of the control unit user interface was found to be important for user efficiency and comfort. The software must not only be functional and enable flexibility, but also has to be designed in a way allowing intuitive use by scientists and all staff not trained in automation. The interface must be easy to understand considering and using common laboratory workstyles such as sequences in multiwell plates. With the liquid handling system used in this study, automation was most beneficial for routine analyses. Adaptation of semi-automated methods to higher or lower analyte concentrations was time-consuming. Extensive operator training was necessary.

The different reproducibility observed for DNA quantification compared to HCP content estimation (Fig. 6) indicate that the remaining manual liquid transfer steps, e.g. of the assay calibration standard, still impact the final results. Systematic differences between the pipetting techniques of operators were reported to cause significant deviations in results [1]. A dilution-dependent positive bias at low dilution factors (up to 1:10) and a negative bias of samples diluted more than 1:100 was reported [13] and also observed in our data (Fig. S4). This might be one reason for the difference of average manual and semi-automated results for DNA measurements. In our case, reduction of manual pipetting in the assay protocol reduced the inter-assay variation of DNA measurements to about half. Thus, reproducibility between operators was improved which is essential if data are generated over longer periods of time by several people. The CHO host cell protein ELISA data showed that semi-automation did not improve precision and reproducibility of one trained operator.

We suggest to implement the proposed methodology in early development where screening of different product variants (lead candidates), materials and/or process conditions is required. Semi-automation allows more flexibility required in this stage of development compared to full automation.

We see potential of semi-automation for any other (spectroscopic) analytical method that requires several sample dilutions and/or aliquot reagent dispense after a defined incubation time, such as UV/Vis absorption- or fluorescence-based quantifications, total protein determination (e.g. Bradford staining), nanoparticle tracking (e.g. NanoSight), or viral titer quantification (e.g. TCID50), to name some examples.

Conclusion and outlook

In the presented study, semi-automation of sample preparation for biochemical analyses reduced operator times and operator-specific influence, thereby increased data consistency, and unburdened staff from repetitive physical tasks such as sample dilution. Critical issues of typical analytical

methods for quality testing such as HCP ELISA, DNA quantification, enzymatic endotoxin detection and SPR-based binding assay were discussed. We showed which challenges might arise in common analytical procedures regarding semi-automation with a simple liquid handling station and how these challenges can be addressed.

For commissioners of analytics, automation bears the potential to reduce personnel cost and on the other hand, reduce variability due to operator influence. Less extensive training of staff might be necessary since the critical steps of sample and standard dilution are carried out by a robot. Procedures should be easily adaptable and (re-)validated. For the concerned staff, automation reduces the risk of hand and shoulder ailments.

A next step to further improve analytical workflows and data quality will be assay miniaturization to reduce sample and reagent volumes, material costs, time and potentially accuracy and reproducibility [24].

To get biopharmaceutical ready for Quality-by-Design approaches as recommended by the authorities, such robust and reproducible analytics can be used to calibrate process analytical technology (PAT) and model-based control algorithms.

Acknowledgements Open access funding provided by the Austrian Centre of Industrial Biotechnology. This work has been supported by the Federal Ministry for Digital and Economic Affairs (bmwd), the Federal Ministry for Transport, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, Government of Lower Austria and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. The funding agencies had no influence on the conduct of this research. We thank our company partners at Boehringer-Ingelheim RCV Process Bioscience and Sandoz/Novartis for their collaboration and fruitful discussions. We also thank Dieter Palmberger for his help in drafting the manuscript and Anna-Carina Frank for technical assistance.

Author contributions AC: conceptualization, data acquisition, data evaluation, manuscript writing and editing. EF: data acquisition, methodology development. MM: data acquisition, assistance in methodology development. DS: development of protein purification and analysis methods, manuscript editing. NW: development of protein purification and analysis methods, manuscript editing. AD: supervision, funding acquisition, manuscript editing. AJ: supervision, funding acquisition, manuscript editing.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest to declare.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. An P, Winters D, Walker KW (2016) Automated high-throughput dense matrix protein folding screen using a liquid handling robot combined with microfluidic capillary electrophoresis. Protein Expr Purif 120:138–147. https://doi.org/10.1016/j.pep.2015.11.015
2. Palandra J, Weller D, Hudson G, Li J, Osgood S, Hudson E, Zhong M, Buchholz L, Cohen LH (2007) Flexible automated approach for quantitative liquid handling of complex biological samples. Anal Biochem 79:8010–8015. https://doi.org/10.1021/ac070618s
3. Yuan L, Zhang D, Aubry A (2012) Automated dried blood spots standard and QC sample preparation using a robotic liquid handler. Bioanalysis 4:2795–2804
4. Liu L, Zhang R, Pajak L (2010) Automation of human IgG, glucose, lactate, and oxygen assays on biomek NXPSpan-8 automation workstation and paradigm detection platform. JALA J Assoc Lab Autom 15:414–418. https://doi.org/10.1016/j.jala.2010.06.010
5. Li M, Chou J, King KW, Jing J, Wei D, Yang L (2015) ICE-CAP: an integrated, general-purpose, automation-assisted IC50/EC50 assay platform. J Lab Autom 20:32–45. https://doi.org/10.1177/2211068214562449
6. Ndumbe PM, Efuetnkeng B, Ikomey GM, Atashili J (2011) Comparison of an automated and a manual method of viral load determination in HIV-infected persons in Yaounde, Cameroon. J Int Assoc Physicians AIDS Care 10:342–344. https://doi.org/10.1177/1545109711398663
7. Antonelli G, Padoan A, Artusi C, Marinova M, Zaninotto M, Plebani M (2016) Automated saliva processing for LC–MS/MS: improving laboratory efficiency in cortisol and cortisone testing. Clin Biochem 49:518–520. https://doi.org/10.1016/j.clinbiochem.2015.12.006
8. Archibald PRT, Chandra A, Thomas D, Chose O, Massouridès E, Laâbi Y, Williams DJ (2016) Comparability of automated human induced pluripotent stem cell culture: a pilot study. Bioprocess Biosyst Eng 39:1847–1858. https://doi.org/10.1007/s00449-016-1659-9
9. Liu S, Boyer-Chatenet L, Lu H, Jiang S (2003) Rapid and automated fluorescence-linked immunosorbent assay for high-throughput screening of HIV-1 fusion inhibitors targeting gp41. J Biomol Screen 8:685–693. https://doi.org/10.1177/1087057103259155
10. Blow N (2008) Lab automation: tales along the road to automation. Nat Methods 5:109–112
11. Whitehead E, Rudolf F, Kaltenbach H, Stelling J (2018) Automated planning enables complex protocols on liquid-handling robots. ACS Synth Biol 7:922–932. https://doi.org/10.1021/acssynbio.8b00021
12. Chan DW (2012) Immunoassay automation: a practical guide. Academic Press, USA
13. Pandya K, Ray CA, Brunner L, Wang J, Lee JW, DeSilva B (2010) Strategies to minimize variability and bias associated with manual pipetting in ligand binding assays to assure data quality of protein therapeutic quantification. J Pharm Biomed Anal 53:623–630. https://doi.org/10.1016/j.jpba.2010.04.025
14. Björksten MG, Almby B, Jansson ES (1994) Hand and shoulder ailments among laboratory technicians using modern plunger-operated pipettes. Appl Ergon 25:88–94
15. Walch N, Scharl T, Felföldi E, Sauer DG, Melcher M, Leisch F, Dürauer A, Jungbauer A (2019) Prediction of the quantity and purity of an antibody capture process in real time. Biotechnol J 4:1800521. https://doi.org/10.1002/biot.201800521

16. Tscheliessnig A, Jungbauer A (2009) High-performance monolith affinity chromatography for fast quantitation of immunoglobulin G. J Chromatogr A 1216:2676–2682. https://doi.org/10.1016/j.chroma.2008.07.063
17. Sauer D, Mosor M, Frank A-C, Weiß F, Christler A, Walch N, Jungbauer A, Dürauer A (2019) A two-step process for capture and purification of human basic fibroblast growth factor from *E. coli* homogenate: yield versus endotoxin clearance. Protein Expr Purif 153:70–82. https://doi.org/10.1016/j.pep.2018.08.009
18. Bland JM, Altman DG (1999) Statistical methods in medical research. Stat Methods Med Res 8:135–160. https://doi.org/10.1177/096228029900800204
19. Jiang H, Ouyang Z, Zeng J, Yuan L, Zheng N, Jemal M, Arnold ME (2012) A user-friendly robotic sample preparation program for fully automated biological sample pipetting and dilution to benefit the regulated bioanalysis. J Lab Autom 17:211–221. https://doi.org/10.1177/2211068211429775
20. Yokota H, Kiyonaga H, Kaniwa H, Shibanuma T (2001) Adsorption of endotoxins on glass in the presence of cationic proteins. J Pharm Biomed Anal 25:1001–1007. https://doi.org/10.1016/S0731-7085(01)00357-0
21. Erridge C, Bennett-guerrero E, Poxton IR (2002) Structure and function of lipopolysaccharides. Microbes Infect 4:837–851
22. Dürauer A, Hobiger S, Walther C, Jungbauer A (2016) Mixing at the microscale: power input in shaken microtiter plates. Biotechnol J 11:1539–1549. https://doi.org/10.1002/biot.201600027
23. Wu JZ, Sinsel EW, Shroyer JF, Warren CM, Welcome DE, Zhao KD, An KN, Buczek FL (2014) Analysis of the musculoskeletal loading of the thumb during pipetting—a pilot study. J Biomech 47:392–399. https://doi.org/10.1016/j.jbiomech.2013.11.015
24. Fernandes P, Carvalho F, Marques MPC (2011) Miniaturization in biotechnology: speeding up the development of bioprocesses. Recent Pat Biotechnol 5:160–173. https://doi.org/10.2174/187220811797579105

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Material

Bioprocess and Biosystems Engineering, Springer

Semi-automation of process analytics reduces operator effect

A. Christler[1], E. Felföldi[1], M. Mosor[1], D. Sauer[1], N. Walch[1], A. Dürauer[1,2], A. Jungbauer[1,2]

[1] Austrian Centre of Industrial Biotechnology, Muthgasse 11, A-1190 Vienna, Austria
[2] Institute of Bioprocess Science and Engineering, Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Muthgasse 18, A-1190 Vienna, Austria
E-mail: alois.jungbauer@boku.ac.at

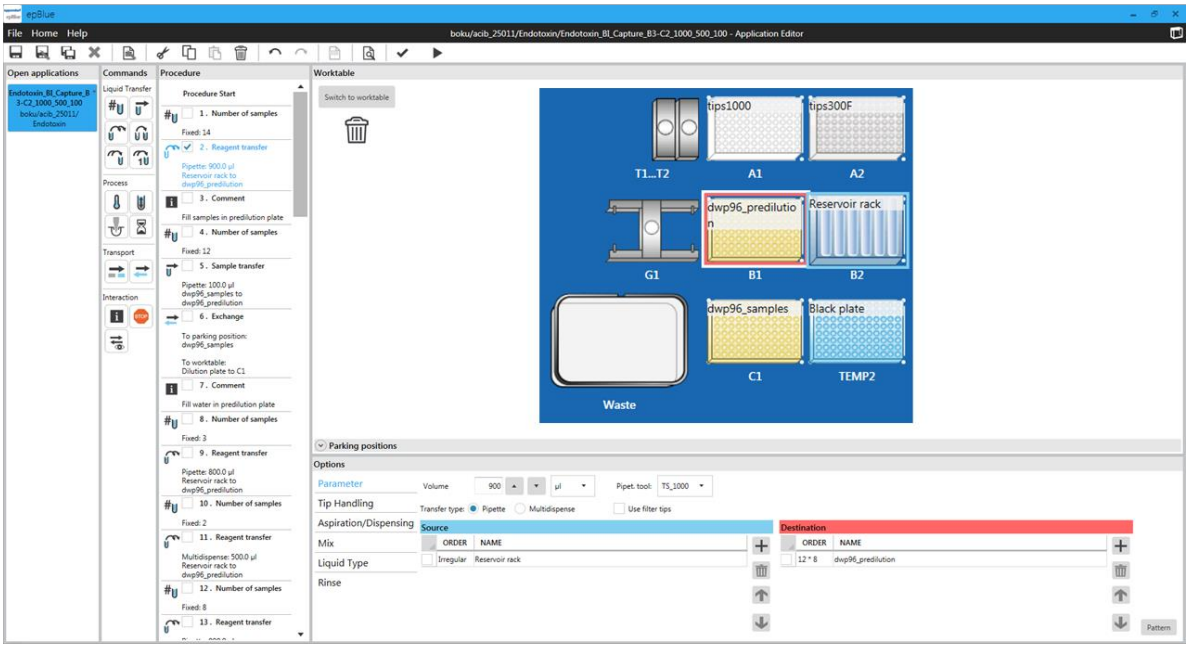


Figure S1: Representation of the epBlue™ Application Editor used to control the LHS.

Calculation and data import into LHS and sample dilution scheme for binding affinity assay

Process samples were diluted in 96-well 2 mL deepwell plates starting in column 1 with a 1:10 dilution. Up to 4 serial 1:10 dilutions were carried out by transfer to columns 2 – 4 (see **Figure S2B**). Higher one-step dilution factors were avoided in all methods for accuracy reasons. In column 5, individual dilutions are performed for each sample to reach a common concentration for all samples (normalization step). In columns 6 to 9 all samples were diluted simultaneously with the 8-channel tool to levels around the dissociation constant K_D of FGF-2 to FGF receptor 2 which was found to be in the range of 1 to 4 nM.

A

Enter concentrations of samples

Location where to take sample for individual dilution

Copy volumes into csv-files

ID	Sample name	Conc. sample (mg/ml)	Conc. sample (nM)	Dil. factor for 20 nM	Dilution (logs)	Logs rounded down	dil. factor remaining	Vol dil. Sample (µl)	Vol. buffer (µl)	Final volume (20 nM)
1	1B1	0.121	7 034.9	351.7	2.5	2	3.5	170.6	429.4	600
2	1B2	0.239	13 895.3	694.8	2.8	2	6.9	86.4	513.6	600
3	1B3	0.688	40 000.0	2 000.0	3.3	3	2.0	300.0	300.0	600
4	1B4	1.946	113 139.5	5 657.0	3.8	3	5.7	106.1	493.9	600
5	1B5	5.471	318 081.4	15 904.1	4.2	4	1.6	377.3	222.7	600
6	1B6	13.631	792 500.0	39 625.0	4.6	4	4.0	151.4	448.6	600
7	1B7	22.889	1 330 755.8	66 537.8	4.8	4	6.7	90.2	509.8	600
8	1B8	23.413	1 361 220.9	68 061.0	4.8	4	6.8	88.2	511.8	600
9	1B9	17.386	1 010 814.0	50 540.7	4.7	4	5.1	118.7	481.3	600
10	1B10	8.948	520 232.6	26 011.6	4.4	4	2.6	230.7	369.3	600
11	1B11	3.544	206 046.5	10 302.3	4.0	4	1.0	582.4	17.6	600
12	1B12	1.093	63 546.5	3 177.3	3.5	3	3.2	188.8	411.2	600
13	1C1	0.419	24 360.5	1 218.0	3.1	3	1.2	492.6	107.4	600
14	1C2	0.277	16 104.7	805.2	2.9	2	8.1	74.5	525.5	600
15	1C3	0.188	10 930.2	546.5	2.7	2	5.5	109.8	490.2	600

B

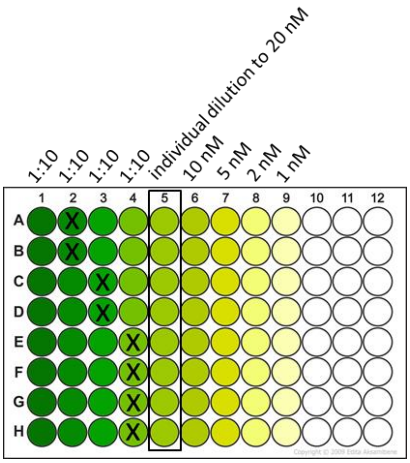


Figure S2: Sample preparation by semi-automated dilution for binding affinity measurement of FGF-2. (A) Calculation sheet to determine dilution steps. Csv-files are loaded into the LHS control software. (B) Scheme of dilution plate (96-well 2 mL deepwell plate). Every sample is diluted to 1, 2, 5, 10 and 20 nM.

Development of the plate wash procedure

A plate wash procedure for HCP ELISA was set-up on the LHS by sequentially dispensing and aspirating wash buffer. In order to remove as much of the wash solution out of the wells as possible in each step, the lowest tip position was determined at which none of the 8 tips touched the bottom of the wells. Keeping a distance to the bottom was important in order not to disturb the bound surface layer. This distance between well bottom and tip was eventually 0.8 mm. The remaining liquid volume after aspiration was 20 – 25 µL/well. In manual plate wash, wells are filled with multi-channel pipette and emptied by gravitational and mechanical force resulting in only the surface adsorbed liquid layer to remain in the wells.

Prior to the washing procedure the well were filled with a GFP solution in order to be able to rapidly and accurately determine the residual amount of protein. After each cycle of liquid aspiration and dispense of fresh wash buffer, the amount of remaining substance was determined by fluorescence measurement of GFP. The signal amplification factor (gain) was increased from 65 to 90 as the amount of substance decreased. Sequences with same number and one additional wash cycle were compared to manual plate wash for their efficiency to remove unbound substances (see Figure 3A). After manual wash the signal of GFP was 9% higher than the blank and in automated wash procedure the difference varied from 9 - 28%. (Figure 3B). The 16% difference compared to 9% was not considered as significant and therefore also included in further evaluations.

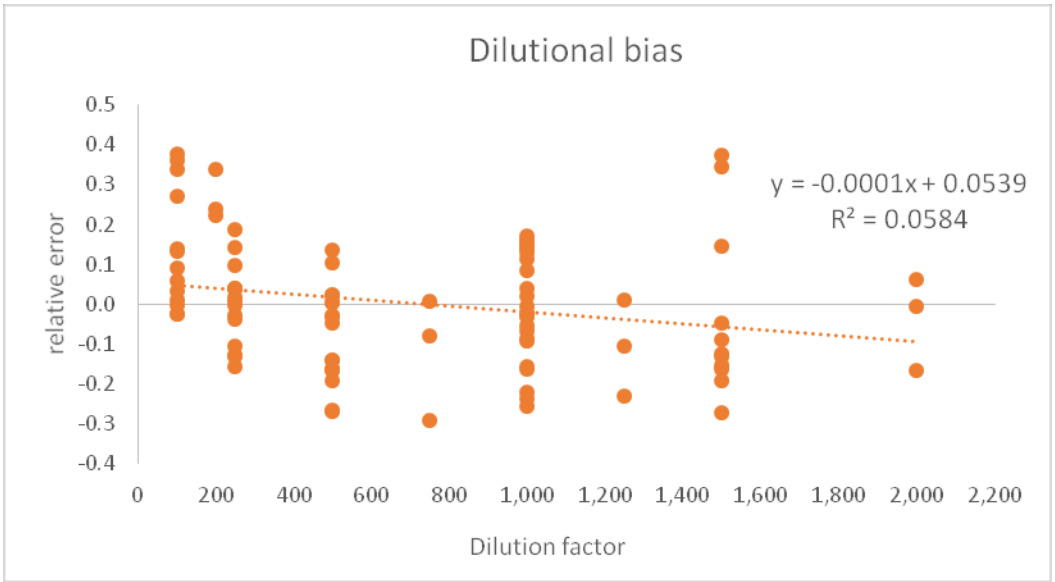


Figure S4: Dependence of results of DNA measurements on dilution factors (proportional bias).

Table S1: Details to Figure 5 on duration of assays in manual and semi-automated ways.

Analyte (Assay)		Same in both methods				Manual in semi-automated			Sample dilution [a]				Total	Working time saved
	Method	Steps	[min]	StdDev [min]	RSD	[min]	StdDev [min]	RSD	manual [min]	Automated [min]	StdDev [min]	RSD	[min]	
Host Cell Proteins (ELISA)	manual	Preparations, plate coating and blocking, 4x plate wash, measuring absorbance, data evaluation, cleaning.	75.8	3.1	4.1%				43.3		0.577	1.3%	119.1	
	semi-automated					31.75	4.99	15.7%		62.7	1.50	2.4%	94.5	11.6
Endotoxins (Factor C)	manual	Preparations, mix reagent and add to samples, measure fluorescence, data evaluation, cleaning.	31.2	4.8	15.2%				55.5		4.5	8.1%	86.7	
	semi-automated					2.3	0.6	24.7%		53.7	0.1	0.2%	56.1	53.2
DNA (Picogreen)	manual		46.2	2.9	6.4%				18.7		1.53	8.2%	64.9	
	semi-automated					4.5	0.2	3.7%		25.6	1.1	4.4%	30.1	14.1
Binding affinity (Biacore)	manual	Calculation of dilutions, start SPR measuring, data evaluation, cleaning.	24	2.0	8%				53.7		2.3	4.3%	77.7	
	semi-automated					21.0	0.8	3.6%		27.6	0.3	1.1%	72.5	32.7
Total (ELISA, Endotoxin assay, Picogreen: 1x, Binding affinity: 2x) [b]									**224.9**	197.1				**144.3**

[a] For ELISA also staining and stop were done manually or automated.

[b] Binding affinity: 8 samples per assay.

Check for updates

Received: 26 February 2021 | Revised: 28 May 2021 | Accepted: 19 June 2021

DOI: 10.1002/bit.27870

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Technology transfer of a monitoring system to predict product concentration and purity of biopharmaceuticals in real-time during chromatographic separation

Anna Christler[1,2] | Theresa Scharl[1,3] | Dominik G. Sauer[1,4] | Johannes Köppl[5] | Anne Tscheließnig[6] | Cabir Toy[6] | Michael Melcher[1,3] | Alois Jungbauer[1,2] | Astrid Dürauer[1,2]

[1]Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria

[2]Department of Biotechnology, Institute of Bioprocess Science and Engineering, University of Natural Resources and Life Sciences, Vienna, Austria

[3]Department of Landscape, Spatial and Infrastructure Sciences, Institute of Statistics, University of Natural Resources and Life Sciences, Vienna, Austria

[4]Biopharmaceutical Process & Product Development, Novartis, Schaftenau, Austria

[5]Technical Operations Large Molecules, Manufacturing Science & Technologies, Novartis, Kundl, Austria

[6]Boehringer Ingelheim Regional Center Vienna GmbH & Co KG, Vienna, Austria

Correspondence
Astrid Dürauer, Department of Biotechnology/ Institute of Bioprocess Science and Engineering, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria.
Email: astrid.duerauer@boku.ac.at

Funding information
Österreichische Forschungsförderungsgesellschaft, Grant/Award Number: 848951

Abstract

Technological developments require the transfer to their location of application to make use of them. We describe the transfer of a real-time monitoring system for lab-scale preparative chromatography to two new sites where it will be used and developed further. Equivalent equipment was used. The capture of a biopharmaceutical model protein, human fibroblast growth factor 2 (FGF-2) was used to evaluate the system transfer. Predictive models for five quality attributes based on partial least squares regression were transferred. Six out of seven online sensors (UV/VIS, pH, conductivity, IR, RI, and MALS) showed comparable signals between the sites while one sensor (fluorescence) showed different signal profiles. A direct transfer of the models for real-time monitoring was not possible, mainly due to differences in sensor signals. Adaptation of the models was necessary. Then, among five prediction models, the prediction errors of the test run at the new sites were on average twice as high as at the training site (model-wise 0.9–5.7 times). Additionally, new prediction models for different products were trained at each new site. These allowed monitoring the critical quality attributes of two new biopharmaceutical products during their purification processes with mean relative deviations between 1% and 33%.

KEYWORDS
critical quality attributes, modeling, online monitoring, PLS, prediction, soft sensor

Anna Christler and Theresa Scharl equally contributing authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Biotechnology and Bioengineering* Published by Wiley Periodicals LLC

1 | INTRODUCTION

According to the WHO, transfer of technology is defined as "a logical procedure that controls the transfer of any process together with its documentation and professional expertise between development and manufacture or between manufacture sites" (World Health Organization, 2011). In this study, we describe the transfer of real-time monitoring and pooling system for preparative chromatographic separation which we developed previously (Sauer, et al., 2019a; Walch et al., 2019). Usually, pooling decisions are made based on off-line or at-line analysis (Mendhe et al., 2015; Rathore, Wood, et al., 2008; Rathore, Yu, et al., 2008; Shekhawat & Rathore, 2019). Our monitoring system allows to predict critical quality attributes such as product concentration and content of process- and product-related impurities by statistical models in real-time. The main advantage of this approach is that it saves a lot of process time compared with the determination of quality attributes by offline wet lab analysis. Moreover, real-time monitoring enables prospective process control (Jiang et al., 2017). Predictive models were built on process data derived from a panel of online sensors and offline analyses of the corresponding quality attributes. As online sensors, conventional pH, conductivity, and UV absorbance measurements were connected in series with commercially available multiangle light scattering (MALS), refractive index (RI), and ATR-FTIR (IR) probes. A prototype fluorescence measurement device was also integrated which allowed to collect emission spectra alternatingly at six different excitation wavelengths. Overall, several thousands of predictors from the online sensors were available for model building. Structured additive regression (STAR) and partial least squares (PLS) regression were applied as modeling techniques (Sauer, et al., 2019a; Walch et al., 2019). In the present work, we used partial least squares (PLS) regression as a modeling technique. PLS regression has been used for similar purposes and with similar online sensors (Grote et al., 2014; Roychoudhury et al., 2006; Wasalathanthri, 2020). This chemometric technique reduces the dimensionality of the data set by projecting the original variables to latent structures. The method is particularly suited for such highly correlated variables generated by the online sensors. For model training, the effluent of the chromatography column was fractionated and analyzed for the respective quality attributes. The optimal number of fractions and replicates of chromatographic runs is a function of the precision of the off-line method and the requested quality of the prediction frequently assessed by the root mean squared error (RMSE) (Felföldi et al., 2020).

The transfer of the monitoring system between different sites is described in this study. A chromatographic capture step of recombinant human fibroblast growth factor 2 (FGF-2) based on ion exchange (Sauer, et al., 2019b) was used as an industry-relevant model process for system comparison. The elution phase was monitored. We used 12 training runs with 15 fractions each to reach a total number of 180 observations. This illustrates that many fractions must be analyzed to establish a model. Hence, a direct transfer of the method from one site to the other would save a lot of time and material. We hypothesize that the transfer of models from the training site to the new sites is possible because sensors of identical configuration of the same vendors where implemented into commercially available chromatographic workstations of the same type at all sites. Furthermore, the same separation process protocol and equivalent feed material were used for all experiments. We tested the hypothesis by evaluating the test errors at the new sites, that is, the difference between predictions and measured quality attributes. The product, FGF-2, was pooled based on offline analyses and based on model-predicted values and the respective models are being described and discussed. Case studies were performed at the new sites showing the functionality with newly generated models specific for the site.

2 | MATERIALS AND METHODS

2.1 | Materials

All chemicals were purchased from Merck unless stated otherwise. Basic human FGF-2 was expressed in *Escherichia coli* BL21 in soluble form. Cells harvest, disintegration, and clarification were carried out as described by Sauer et al. (2019b). Aliquots of the homogenates obtained from fermentation broth carried out under same conditions were used as feed material for the experiments at the three sites. For the case studies, biopharmaceutical proteins were produced by proprietary processes.

2.2 | Methods

2.2.1 | Chromatography

FGF-2 was purified by chromatographic purification on an Äkta Pure 25 (Cytiva) as described in Sauer, et al. (2019b). In brief, *E. coli* homogenate was 0.22 μm filtered and 118 ml (10 CV) of the clarified homogenate were loaded on a column packed with weak cation exchanger Carboxymethyl-Sepharose Fast Flow (Cytiva) with 11.8 ml CV (1 cm diameter, 15 cm bed height). The column was equilibrated before and washed after loading with 100 mM Na-phosphate, pH 7.0. FGF-2 was eluted by a linear gradient from 0 to 1 M NaCl in 100 mM Na-phosphate pH 7.0. During the elution phase, the effluent was collected in 1 ml fractions. Fifteen fractions were analyzed around the peak center. The column was sanitized after each run with 1.0 M NaOH for 1 h (5 CV). For the case studies, proprietary purification protocols were used. For model training, 8–9 replicate runs were performed at each of the new sites, 6–7 of them as training runs, and 2 runs as test set.

2.2.2 | Online sensors and database

Sensors were integrated in the column effluent stream in-line in the order of increasing flow cell void volume and/or increasing pressure sensitivity. Details are described in Sauer, et al. (2019a) and Walch et al. (2019). A mid-infrared spectrometer MATRIX-FM (Bruker) was used to measure ATR-FTIR spectra in the wavenumber range from 3500 to 750 cm^{-1} at a resolution of 2 cm^{-1}. Intrinsic protein

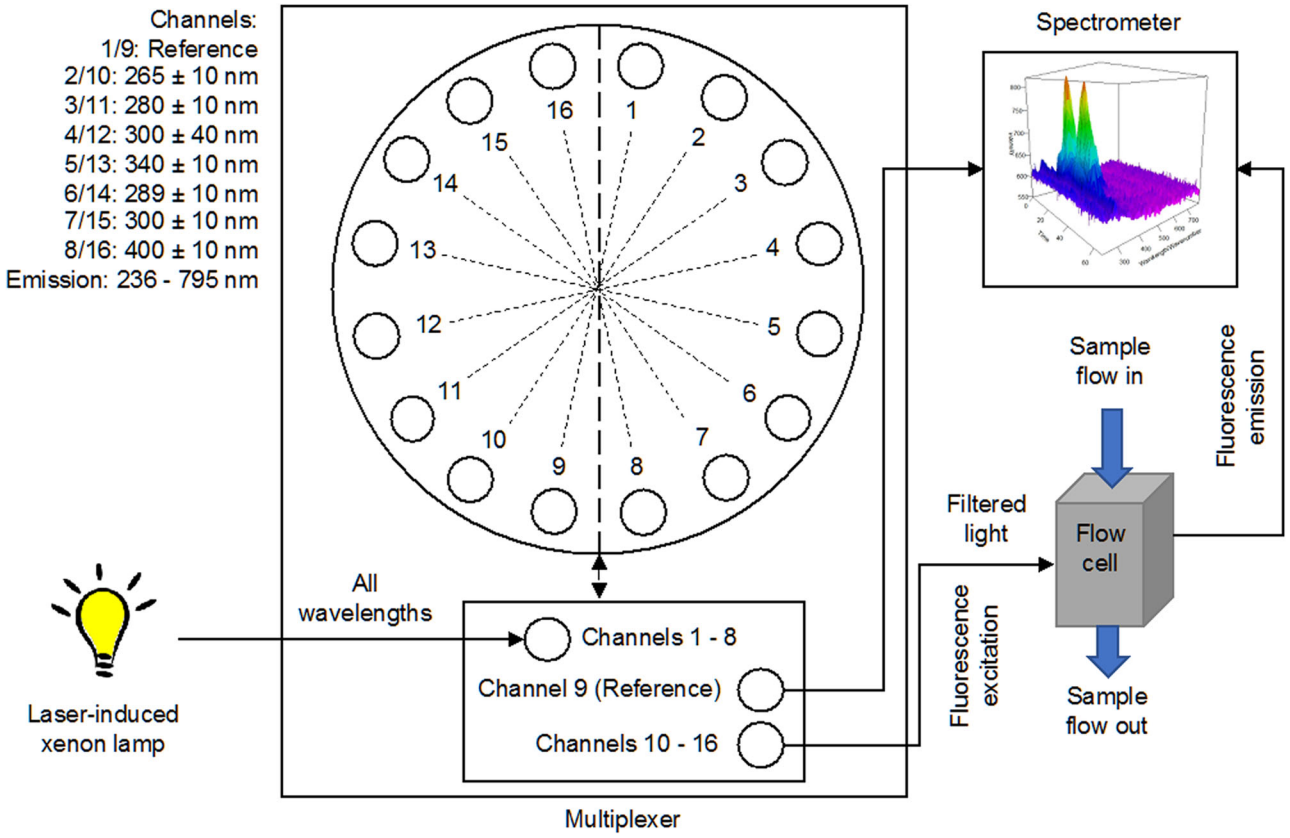


FIGURE 1 Schematic of the in-house assembled multi-wavelength fluorescence detector. Arrows represent optical fibers. Not drawn to scale

fluorescence was recorded at emission wavelengths between 236 and 795 nm at a resolution of 0.3 nm. Excitation was done at six different wavelengths and one (300 nm) with a small and a large filter width: 265 ± 10 nm, 280 ± 10 nm, 289 ± 10 nm, 300 ± 10 nm, 300 ± 40 nm, 340 ± 10 nm, and 400 ± 10 nm. The sensor was assembled in-house using a laser-induced xenon lamp (type EQ-99XFC LDLS, Energetiq), a fiber optic multiplexer (Avantes), a flow cell (FIAlab Instruments), and the spectrometer AvaSpec-ULS-TEC (Avantes) (Figure 1). Measurement time for all seven emission spectra including multiplexer switching time was 16 s. All other detectors were standalone commercial devices. An RI detector Optilab T-rEX (Wyatt) was used allowing differential RI detection in the range of −0.0047 to + 0.0047 RIU. The RI detector also traced a forward monitor for evaluation of system integrity and the LED intensity. Light scattering signals were recorded with the MALS detector miniDAWN TREOS (Wyatt) at angles of 43.6° (LS1), 90° (LS2), and 136.4° (LS3). Additionally, forward monitor intensity and temperature were recorded.

Volume delay between sensors was determined gravimetrically and sensor signals aligned accordingly before modeling and prediction. All buffers used in the process were aqueous based, therefore water was used as blank for all measurements. Blanks were measured with UV, IR, RI, and MALS before each run and signals adjusted. The pH probe was calibrated with a linear calibration between pH 4 and pH 7 before each run. The IR detector was cooled with liquid nitrogen at least 20 min before each run. Signals were recorded and stored by the control software XAMIris (evon).

2.2.3 | Offline analytics for quality parameters

All offline analyses were performed as described in Sauer, et al. (2019b). In essence, protein quantity was determined by reversed-phase (RP) HPLC using a TSKgel Super-Octyl column (2 μm bead diameter, 4.6 × 50/100 mm, 110 Å, Tosoh Bioscience). Monomer and high molecular weight impurity (HMWI) contents were determined by their relative peak areas after size-exclusion (SEC) UPLC using an ACQUITY UPLC BEH125 SEC column (1.7 μm bead diameter, 4.6 × 150 mm, Waters). Low molecular weight impurities were present in negligible amounts in the peak center.

Host cell proteins (HCP) were determined by anti-*E. coli*-HCP sandwich ELISA in 96-well plate format with antibodies from Cygnus. Values of 3–6 doubling dilutions per sample were averaged. Double-stranded DNA (dsDNA) was quantified by Quant-iT® Picogreen (Thermo Fisher Scientific) fluorescence dying in 96-well plate format. Values of 3–4 doubling dilutions per sample were averaged. Offline analyses for the case studies were performed by proprietary analytical methods.

2.3 | Statistical modeling

PLS regression was used to generate prediction models of quantity and purity. All data handling was performed within the statistical computing environment R (Team R. C., 2020) using R package pls (Mevik et al., 2019) as in Walch et al. (2019). This linear regression method is particularly suited for multicollinear variables. Models were trained on 8–12 replicate runs (120–180 data points) at the training site as there were some missing sensor data at the training site. Targets were the quality attributes quantity (measured in mg/ml), monomer content (in %), HMWI content (in %), and HCP and dsDNA contents (both in ng/ml). Based on expert knowledge (such as amide bands and fingerprint regions), the spectral data were reduced to certain regions of interest and different combinations of sensor signals were tested. Numerous subsets of in total 15,725 predictors (online variables) that were available and usable at all three sites were selected. Prediction models were generated on the training data using leave-one-run-out cross-validation (CV) for each of the five responses. The error measure used here was the RMSE, which is a measure of the average prediction error. It is given in the unit of the respective quality attribute and calculated by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

where n is the number of samples, y_i are the measured values, and $\hat{y}_i$ the predicted values. The set of predictors yielding the lowest RMSE for a given response was selected for the prediction model of this response. In case of equivalence or almost equivalence between models for one target, models with fewer predictors were preferred for reasons of simplification and greater robustness against sensor fall-outs. Prediction quality was assessed by applying the models on test data sets which were not used for model training. The error measure is then called "test RMSE."

To assess the quality of the proposed models we also included so-called null models. In a null model the response of a certain run is simply predicted by the average values of all the training runs without the use of any predictors. If the proposed model outperforms the null model, the contained predictors are considered important and useful for the prediction of the response.

Another measure of the quality of prediction used in this study was the mean relative deviation (MRD, in %) defined by

$$\mathrm{MRD} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|.$$

Contrary to the RMSE, the MRD is a relative measure and can be regarded as independent of the range of the measured values. Therefore, it was used for the case studies where not all information could be disclosed.

2.4 | Method transfer to new sites

Online sensors at two new sites (A and B) were purchased from the vendors mentioned above. Offline analyses were performed at the training site to reduce differences between sites. Operators received hands-on training for the use of the monitoring system. A run checklist was transferred to ensure correct system handling. Three multi-day hands-on trainings were performed for data analysis and modeling.

Data of runs B1 and B2 were shifted forward by 2 and 1 min respectively, corresponding to equal milliliters, for easier visual comparison.

2.5 | Pooling

Fractions were manually selected for pooling so that the pools met defined quality criteria while maximizing the yield (collected protein over eluted protein). Negative predictions of dsDNA, HCP, and HMWI were set to 0 before fraction selection for calculation of the pool averages. Predictions of more than 100% Monomer were set to 100%. For pooling, HCP and dsDNA were calculated in parts per million (ppm) by dividing the values in ng/ml by the FGF-2 concentration in mg/ml. Runs were pooled independently based on measured quality attributes and based on model predictions. The pooled fractions can be identified in Online Supporting Information.

2.6 | Case studies

All offline analyses were performed at the new sites. PLS models for six critical quality attributes were trained on 6–7 training runs and tested on data of two independent test runs. In each run, 13 fractions were collected.

3 | RESULTS

Online monitoring systems for chromatographic purification were set-up at the two new sites. Automated sensor control was enabled by custom control software. The functionality of each system was first tested using a human serum albumin solution (data not shown). As a model process, we used the chromatographic capture process of FGF-2 from clarified *E. coli* homogenate. First, models were established for real-time prediction of product concentration and contents of monomer, HMWI, HCP, and dsDNA at the training site. These models were transferred to the new sites.

3.1 | Online sensor data of FGF-2 capture at three sites

Online sensor data of three FGF-2 capture runs at site A (runs A1–A3) and three runs at site B (runs B1–B3) were compared with the arithmetic mean of 12 runs performed at the training site (Figure 2). To include the variability at the training site, one, two, and three standard deviations (*SD*) of the signals obtained at the training

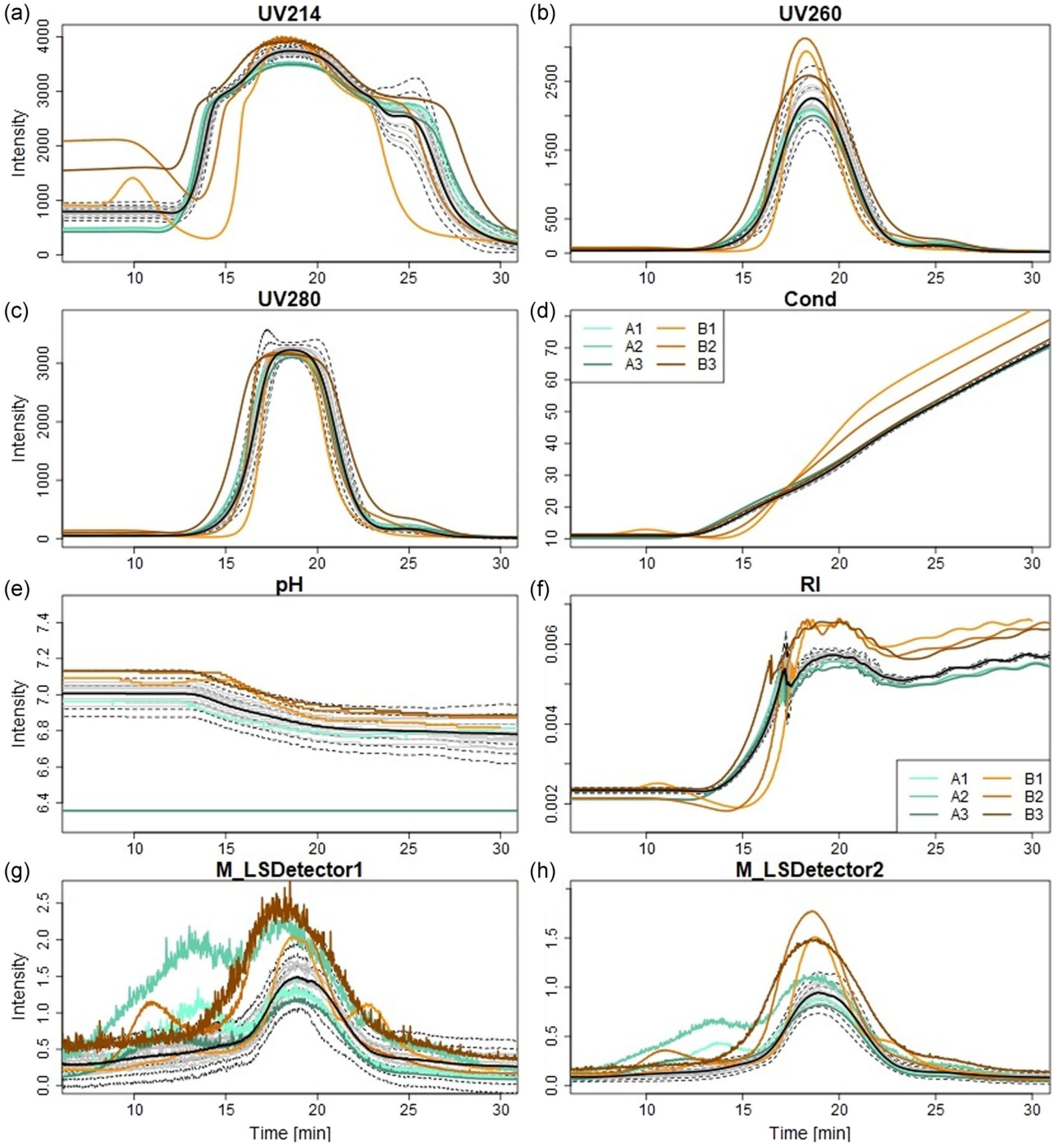


FIGURE 2 Online sensor data during the elution phase from three test runs at sites A and B compared with data from the training site (gray lines). Black lines: averages of 12 training runs. Dashed lines: 1, 2, and 3 standard deviations of training data. Sensors: UV (214, 260, and 280 nm, a–c), conductivity (d), pH (e), refractive index (f), light scattering (g) and (h)

site were included in the graph. The three runs and the respective sensor signals at site A were highly reproducible except for the MALS detector. At site B, the runs were not as reproducible. Elution was delayed in runs B1 and B2 as can be seen in the UV absorption, conductivity, and refractive index signals (Figures 2a–d and 2f). The reason for this delay was a later start and then steeper salt gradient used for protein elution due to insufficient priming of the tubes before the runs. As a result, peaks were narrower with higher peak protein concentrations in runs B1 and B2 (Figure 2b). All sensors except conductivity and pH showed higher maximum intensities compared with site A and the training site. Modifications of the process conditions of the final run B3 caused a broadening of the peak. UV absorption signals showed large differences at site B compared to the other two sites where the signals were comparable

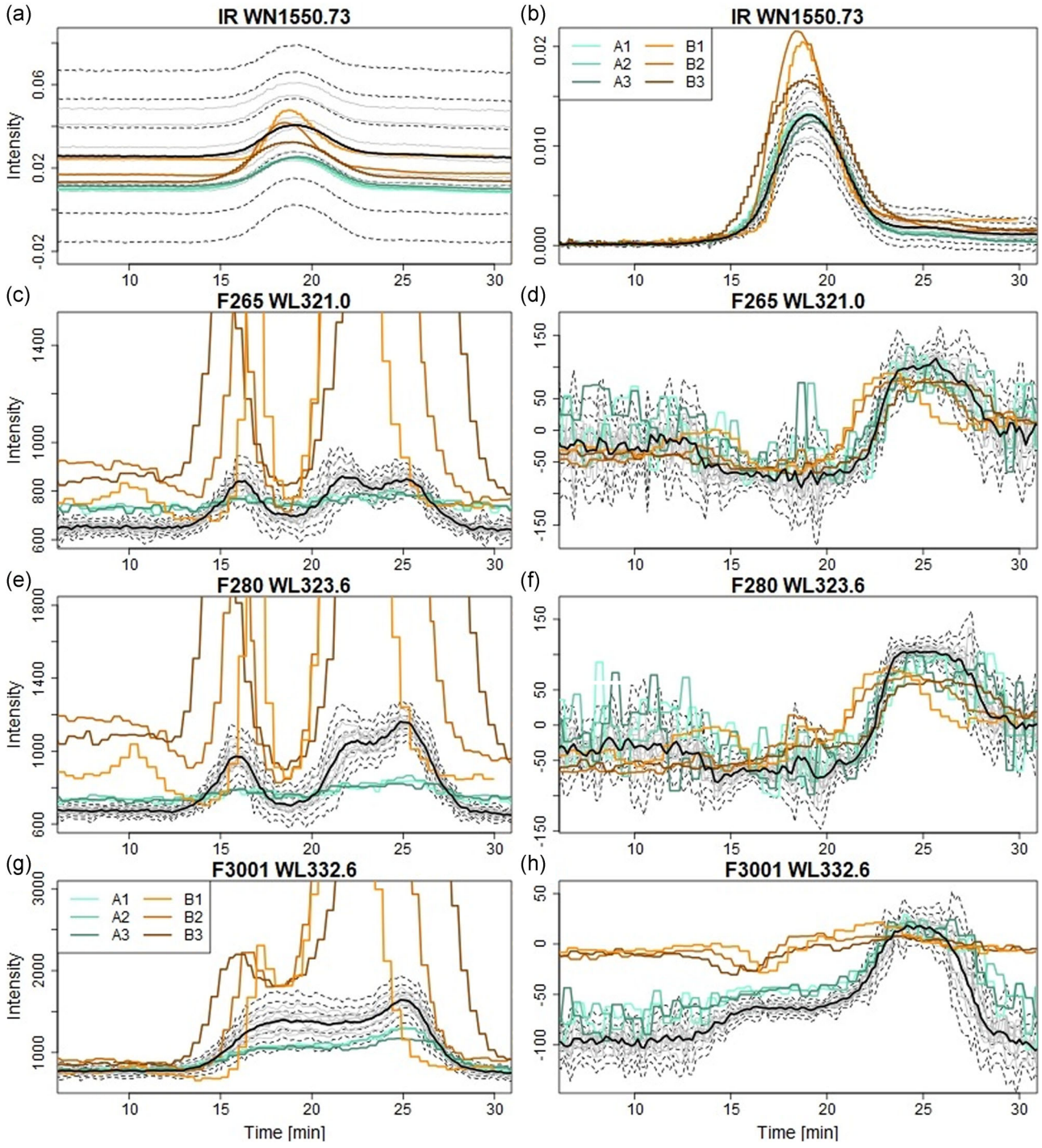


FIGURE 3 Exemplary wavenumbers and wavelengths of IR (a) and (b) and fluorescence data (c)–(h) before and after preprocessing (left and right column, respectively). Gray lines: model training data. Black lines: averages of 13 training runs. Dashed lines: 1, 2, and 3 standard deviations of training data

(Figure 2a–c). pH was not available for runs A2 and A3, but similar in all other runs at all sites (Figure 2d). More pronounced pre-peaks were observed in MALS signals at the new sites in the range of 5–15 min after start of the elution phase (Figure 2g,h). Noisy scattering signals at 43.6° angle (Figure 2g) indicated the requirement for cleaning of the flow cell at both new sites. Processing of cell homogenates led to (fast) fouling of the flow cell. Light scattering at 90° (Figure 2h) was less affected by fouling. Overall, the signals from the two new sites often differed significantly from the training site.

IR and fluorescence sensors recorded absorption and emission spectra, respectively, at each time point in a very high resolution. Thus, representative wavenumbers and wavelengths in the center of the peaks were visualized over the time of the elution phase. Figure 3 shows raw and preprocessed IR and fluorescence data.

Spectral data needed one or more of the following preprocessing techniques: smoothing, baseline correction, first or second derivative, and normalization to equal length or area. Preprocessing methods and their parameter values influenced the suitability for predictive modeling (not shown). They must be optimized iteratively by comparing a score such as the RMSE. Here, smoothing, taking the first derivative and normalization to equal length were selected for fluorescence. IR was preprocessed by subtraction of a reference spectrum recorded before elution in addition to smoothing and baseline correction by asymmetric least squares (Eilers & Boelens, 2005) implemented in the R package *baseline* (Liland et al., 2010).

Among all sensors, the fluorescence data showed the largest differences between sites. Intensities of peaks were different at all sites (Figures 3c, 3e, and 3g). For excitation at 265 nm and 280 nm, signal preprocessing by smoothing, derivatization (first derivative), and normalization to equal length led to a higher agreement of the spectra (Figure 3c–f). For all higher excitation wavelengths (289–340 nm), no suitable preprocessing method or a combination thereof could be found to compensate for the differences of the signals obtained at the different sites (e.g., Figure 3g,h). In addition, a decrease of signal intensities over time as a result of the aging of the lamp and optical fibers was observed (see Figure S1).

3.2 | Model predictions

Predictive models for all described quality attributes were applied to the online data obtained at the new sites. Due to the large differences in fluorescence signals which could not be compensated by preprocessing, the RMSEs of the models were large in the beginning. Fluorescence variables were only included for excitation at 265 and 280 nm. Models for monomer, HMWI, and HCP which contained fluorescence excitation variables above 280 nm, were retrained with data from the training site to be able to apply them to the new sites' data. Figure 4 shows the results of the offline analyses obtained for FGF-2 concentration and contents of monomer, HMWI, HCP, and dsDNA. Furthermore, corresponding model predictions and differences between measured and predicted values (errors) are shown for

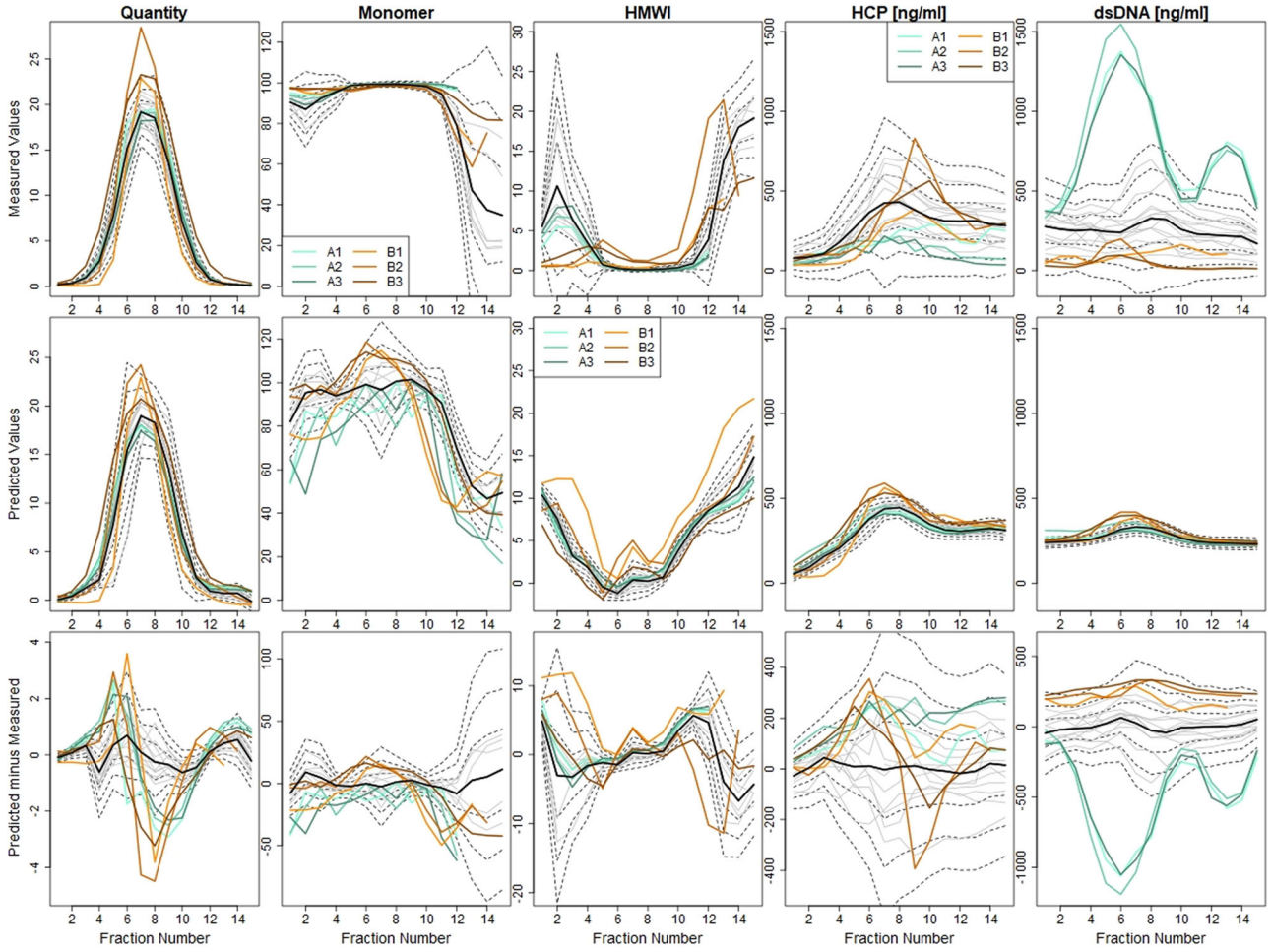


FIGURE 4 Fraction-wise offline measured quality attributes for FGF-2 concentration, monomer, HMWI, HCP, and dsDNA content, model-based predictions, and corresponding differences to offline measured values (prediction errors). Gray lines: model training data. Blackline: average of training data. Dashed lines: 1, 2, and 3 standard deviations of training data. dsDNA, double-stranded DNA; FGF-2, fibroblast growth factor 2; HCP, host cell protein; HMWI, high molecular weight impurities

TABLE 1 Predictors used in the models, RMSEs of the respective null models, and PLS models for test runs at three sites

Model	Predictors (number)	RMSE	Training site	Site A	Site B
Concentration (mg/ml)	UV, RI, conductivity (5)	Null model	7.0	7.0	8.8
		PLS model	0.8 (0.5–1.0)	1.2 (1.1–1.3)	1.1 (0.9–1.4)
Monomer (%)	UV, RI, Fluor265 (25)	Null model	21.9	11.1	24.9
		PLS model	8.5 (4.7–10.4)	17.7 (14.4–20.9)	16.7 (14.9–19.2)
HMWI (%)	UV, conductivity (4)	Null model	6.2	3.7	4.9
		PLS model	3.2 (2.0–4.7)	2.8 (2.7–3.0)	4.2 (1.8–6.1)
HCP (ng/ml)	UV, conductivity, MALS (8)	Null model	156	169	191
		PLS model	81 (37–125)	171 (113–209)	121 (98–141)
dsDNA (ng/ml)	MALS (4)	Null model	110	623	199
		PLS model	87 (39–139)	496 (488–501)	233 (186–271)

Note: The ranges of test RMSEs per run are given in brackets.

Abbreviations: dsDNA, double-stranded DNA; FGF-2, fibroblast growth factor 2; HCP, host cell protein; HMWI, high molecular weight impurities; MALS, multiangle light scattering; PLS, partial least squares; RI, refractive index; RMSEs, root mean squared errors.

TABLE 2 Quality attributes of the collected FGF-2 product pools at the three sites based on model-predicted values and based on offline measurements, respectively

	Training site (n = 6)		Site A (Run A2)		Site B (Run B3)	
	Measured	Predicted	Measured	Predicted	Measured	Predicted
Pool volume (ml)*	8 or 9	8 or 9	4	5	10	10
FGF-2 concentration (mg/ml)	11.8 ± 0.6	10.5 ± 0.4	15.2	13.9	12.5	11.7
Monomer (%)	96.0 ± 1.5	97.6 ± 1.6	99.5	91.1	97.1	93.7
HMWI (%)	0.8 ± 0.4	1.7 ± 0.2	0.1	1.4	2.5	2.0
HCP (ppm)	44 ± 6	56 ± 3	12	34	45	53
dsDNA (ppm)	34 ± 14	56 ± 5	56	26	4	46
Yield (%)	97.1 ± 0.3	96.3 ± 1.3	65.1	73.0	97.6	95.9

Note: Plus/minus values represent one standard deviation. For the new sites, data of one representative run is shown.

Abbreviations: dsDNA, double-stranded DNA; FGF-2, fibroblast growth factor 2; HCP, host cell protein; HMWI, high molecular weight impurities.

*For the exact fractions refer to Online Supporting Information.

each fraction. As before, data from new sites were plotted over the data obtained at the training site including their respective averages and 1, 2, and 3 standard deviations thereof. Predictors selected in each model are given in Table 1.

Errors of the prediction of concentration were distributed around zero in a slightly larger range as for the training data with errors up to 4.5 mg/ml. The monomer content was underpredicted at site A by −8.9% on average and by up to −20.5% in collected fractions. The monomer content was over- and underpredicted at site B by up to 31.2% overall and up to −20.7% in collected fractions. On average, the monomer content at site B was underpredicted by −1.5% in collected fractions. HMWI showed lower absolute values in the beginning of the peak at site B compared with the other sites which was not recognized by the model. Errors of the HMWI predictions were in the same range as at the training site (±12% HMWI). The model for HCP was not able to recognize the different profiles at the new sites, yet the errors were in a similar range as for the training runs. Measured dsDNA was higher at site A compared to the training site and lower at site B, due to the different fermentation batches. The model for dsDNA was not able to predict the different dsDNA profiles accurately. Average prediction errors for dsDNA were about two and five times as high as at the training site for site A and B, respectively (Table 1).

Model quality can also be evaluated with regard to the errors obtained by so-called null models which predict the target value by simply taking the average of the respective variable from all the

training runs. Any trained model has to outperform the null model. For the training site, data from six independent test runs from two new fermentation batches (3 runs per batch) was used as a reference (Sauer, et al., 2019a). These test runs represent a similar situation as was faced at the new sites since the data was not used for model training and the processed material was from different fermentation batches. Predictions and offline measured values can be found in Online Supporting Information.

The PLS model for FGF-2 concentration performed much better than the corresponding null model at all three sites (Table 1). The model for monomer content also performed better than the null model at two of three sites. At site A, the PLS model gave a higher RMSE than the null model. This was probably due to many missing observations at the tail of the peak where the error usually was the largest (compare Figure 4). At the training site, the models predicting HMWI, dsDNA, and HCP allowed better predictions compared with the null models. Average RSMEs of the null models were between 1.3 and 8.8 times larger than the RMSEs of the PLS models at the training site. At the new sites, the advantage of the models over the null models was moderate or none at all. In three cases (monomer at site A, HCP at site A, and dsDNA at site B) the model prediction errors were even higher than the null model RMSE. The performance of the transferred PLS models at the new sites was only satisfying for the FGF-2 concentration.

3.3 | Product pooling based on model predictions

The ability of the new systems to produce product that meets specified quality criteria was evaluated by pooling fractions of each of the test runs. Quality attributes of pools based on model-predicted values were compared to pools based on offline measured data. The following criteria had to be fulfilled by the pools: FGF-2 monomer more than or equal to 90%, HMWI less than or equal to 5%, HCP less than or equal to 60 ppm, DNA less than or equal to 60 ppm, and FGF-2 concentration in the fraction of at least 1.0 mg/ml. Measured and predicted HCP and dsDNA contents in ng/ml were converted into ppm by division through the measured and predicted FGF-2 concentrations, respectively. In an iterative process, fractions were included to maximize the share of collected product from the total eluted protein (=yield) while fulfilling all quality criteria. Table 2 shows the obtained pool quality attributes at the training site and for exemplary runs at the new sites. Run A2 was selected randomly since all runs at site A were very similar. Run B3 was selected as an example because in this run the process conditions best matched those of the other sites. Pool volumes were multiples of 1 ml fractions and thus integer numbers. At the training site, very similar pools were obtained using the predictions and the offline measurements. At site A, less fractions were pooled offline than online due to high measured dsDNA contents. At site B, the model predictions led to the collection of the same fractions as were pooled offline. The FGF-2 concentration in the peak center was underpredicted about equally at all sites: On average, the collected FGF-2 mass was predicted 7.0% lower than the measured mass. This was most probably due to the saturation of many sensors at the peak center, such as UV or RI detectors. Despite deviations in the monomer prediction at the training site of up to 17% in collected fractions, the average monomer content of the pools was similar to the reference analytics. At site A, the monomer content was the limiting variable for pooling based on predictions due to its strong underprediction which led to moderate yields of 73%–88%. The mass balance of %monomer and % HMWI was not closed either due to the presence of low molecular weight impurities (not shown) or due to inaccuracies of the predictions. Mass balances were not considered with these models.

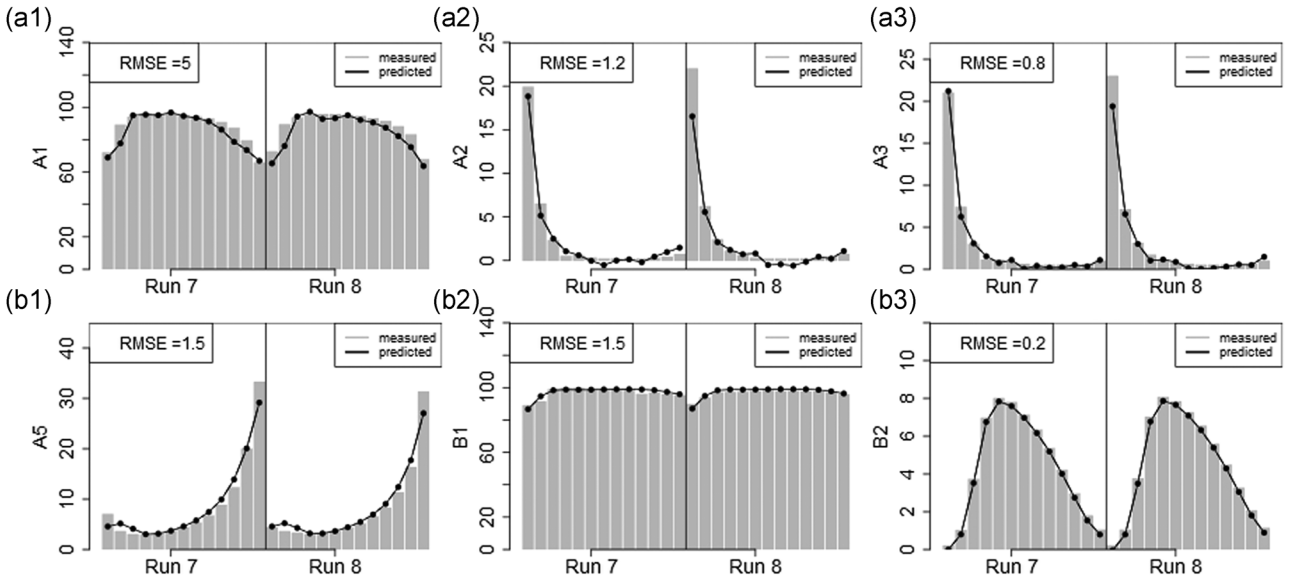


FIGURE 5 PLS predictions and offline measurements of six critical quality attributes of two test runs at one of the new sites. The test RMSE is given for each attribute. PLS, partial least squares; RMSEs, root mean squared errors

The high dsDNA content of the material at site A (compare Figure 4) was responsible for the low yields of the pools based on offline measurements of 65%–82%.

3.4 | Case studies at the new sites

Case studies at the new sites were performed to test the abilities of the systems to predict critical quality attributes of biopharmaceutical proteins during chromatographic separation. One of them is exemplarily shown in Figure 5. Among the modeled product attributes were typical process-related and product-related quality attributes, respectively. Mean relative deviations (MRDs) among the six models were between 1% and 33%.

Among the critical quality attributes were charge variants which represent a form of micro-heterogeneity. The composition of charge variants is a major quality attribute that needs to be controlled in the manufacturing of biopharmaceuticals (Hintersteiner, Lingg, Janzek, et al., 2016; Hintersteiner, Lingg, Zhang, et al., 2016). Protein charge variants have very similar structural and spectroscopic properties and present a challenge for online monitoring and prediction. The developed models were able to predict the different critical quality attributes with satisfying accuracy (not shown).

4 | DISCUSSION

The aim of the technology transfer was to ensure that the systems are fit for their purpose at the new sites. The performances of the online monitoring systems transferred to the new sites were compared with the training site by two different measures: the RMSEs of the test runs quantified the average errors overall fractions and thus the overall quality of the PLS models. The pooling example illustrated the ability of the models to produce qualitative products.

Even though equivalent load material was processed following the identical protocol and using online sensors from the same suppliers at all sites, significant differences between the sensor signals were observed for the system established at the two new sites. Site-specific influences such as operator effects, ambient temperature, humidity, or chemicals might have influenced the processes. Fouling is another common problem when working with biological solutions. Usually, signals are being background corrected by resetting them just before the measurement however at the cost of decreased sensitivity with an increasing amount of fouling. Fouling and sensor aging must be continuously monitored and controlled to ensure that the monitoring system is in a functional state. The IR detector came with a built-in performance qualification test which was done before each run. The test fails, for example, when the device is not cooled enough, when the lamp intensity is too low or humidity too high which would indicate leakage. IR spectra were strongly influenced by ambient conditions, for example, temperature. IR absorption measurements are usually background corrected by a blank spectrum recorded just before the measurement. This was not sufficient here due to the long duration of the loading process of about 2 h. Spectra of different runs showed very different appearances. Subtraction of a spectrum just before the elution was necessary to use the sensor data for modeling. For RI and MALS detectors, Forward monitor intensity of more than 90% was used for quality control. With these measures, the commercial sensors were robust and delivered comparable results across the sites. The reason for the different fluorescence signals was probably that this sensor was not an optimized commercial setup but an in-house assembled prototype (see Figure 1). The flow cell was free-standing and not encased as in the other sensors. The optical fibers must be manually arrested and the exact position impacted the measurement through the transmission of light. Mechanical switching of channels by the multiplexer must be precise to transmit all light. This setup was chosen to allow the scanning of six excitation wavelengths in parallel to gather as much information as possible. However, this was at the expense of the robustness of this fluorescence sensor. In a manufacturing environment, a simpler and more robust sensor would be needed. Fluorescence signals at site A were closer to the training site than signals at site B. The technology was transferred to site A about 2 years after set-up at the training site was completed and to site B about 1 year after that, that is, 3 years after set-up at the training site. Signal preprocessing could not level out the differences sufficiently. Lamp aging is generally common for such lamps and can be predicted but here the effects of fiber aging and lamp aging overlayed.

Accurate representative process data is a major requirement for the generation of reliable prediction models. Therefore, sufficient time and resources must be invested in data generation. However, data generation was work-intensive (compare Christler et al., 2020). The appropriate number of fractions and number of runs depends on the coefficient of variation of the reference analytics (Felfödi et al., 2020). If more fractions are analyzed, less runs can be sufficient because a certain number of total data points is necessary. More observations usually increase the reliability of predictions. More than the planned runs were necessary at the training site because at least one sensor showed fouling, aging, or had a defect. Moreover, the complex system required experience to handle it correctly. Offline and online data of the runs at site A were generally more similar to the training site than at site B. This shows that technology transfer concerns not only the transferred technology but also the technology handling process. System handling is expected to become much easier for an optimized commercial sensor battery.

Model training was conducted within a few days, once the workflow was established and knowledge was obtained on data processing. The main parameters for model optimization were data preprocessing techniques and predictor selection. Simpler models with fewer predictors were selected in this investigation whenever the RMSEs were similar between models containing a different number of predictors. Simpler models bear a lower risk of dysfunction of the monitoring system if single sensors become dysfunctional or lose their connection to the database during a run. Missing predictors can lead to totally missing predictions because all predictors of a model are needed to calculate the target. Missing predictors

could be imputed by other highly correlated predictors using for example tree-based modeling techniques (Kuhn & Johnson, 2013), however at the cost of increased model complexity.

Overall, the UV absorbance sensor gave the most useful predictors, conductivity the second most useful, then RI and MALS, and last fluorescence. IR was not used in a model. However, this is no general statement. In our previous work, fluorescence and infrared sensors yielded very useful predictors, especially for process-derived impurities such as HCP and DNA, but also for monomer and HWMI (Sauer, et al., 2019a; Walch et al., 2019). This could be a reason for the poorer performance of these models here, since fluorescence from excitation above 280 nm could not be used. The fact that a sensor is not included in a model does not necessarily mean that it is useless in predicting a response. In the case of several prediction models with a similar performance the simpler models were selected due to higher robustness. RSMEs of the five models on the test runs of the new sites were on average twice as high (2.0 times) as the test errors at the training site (0.9–5.7 times). Higher errors at the new sites were expected since some of the new data were not represented in the training data set, such as the high dsDNA values at site A or the high HCP values at site B. In such cases, predictions may not be accurate (Kuhn & Johnson, 2013). Variations in product and impurity contents are common in biological processes. The training and test data set should include at least as much variation as is expected later in the application, better a bit more.

Product pooling based on model predictions was possible at all three sites with yields between 73% and 99%. Nevertheless, the differences between offline measurements and model predictions showed that the accuracy of the transferred models was not sufficient for monitoring in the manufacture of biopharmaceuticals. A properly functioning fluorescence sensor is expected to significantly improve the models' performance. Moreover, the test situation was rather complex due to the very different impurity contents between the sites.

The case studies showed that on-site model training allowed to predict six critical quality attributes with good accuracy. A mean relative deviation of 33% may be borderline for manufacturing but there is still room for optimization, for example by using different modeling techniques. Nonlinear techniques such as STAR used in our previous work (Sauer, et al., 2019a) generally bear an advantage in cases where predictors and responses are nonlinearly related. However, nonlinear modeling techniques often need more computation power and time for training. An advantage of PLS is that it can be trained very fast. For proof of principle, we used the simpler method here.

A prerequisite for real-time process monitoring and control is a central powerful data collection point. Large data was generated: for one run about 450 MB of process monitoring data. Sufficient computing capacity was necessary to allow the estimation of several quality parameters within a few seconds. Overall, time for quality estimation could be reduced by days because no offline analytics was necessary anymore. Once the functionality of the monitoring system was shown, it can also be used for process development (e.g., Chemalil, 2020). Sensor data can be used directly, for example, the fluorescence pattern to distinguish protein of interest from HCP. Model predictions can be used to compare process variants as long as the conditions are within the trained design space of the models.

In summary, we were able to transfer statistical models for real-time prediction of five critical quality attributes of a recombinant human FGF-2 process from the training site to two other sites. Up to 5.7 times higher test errors at the new sites compared with the training site were observed with an average of 2.0 times among all quality attributes. The accuracy of the transferred models would not be high enough for real-time process monitoring and product pooling. Model re-training would be needed for application in manufacturing. However, these were not optimized models but used to assess system similarities and general functionality under real process conditions.

5 | CONCLUSION

For a successful transfer of a statistical model for real-time prediction we conclude that the quality of prediction at new sites depend on (a) how close the process parameters can be matched with the training site and (b) how robust and reliably the sensors work at the different sites. The biggest source of errors in our work was the different sensitivity of these highly sensitive sensors. Differences between sensor signals at the three sites could partially be compensated by preprocessing methods and this is considered as advantage of the statistical models. For process monitoring, model re-training at each site was necessary. The case studies showed that on-site model training allowed to predict six critical quality attributes with good accuracy. We conclude that it will be necessary to at least optimize transferred statistical models at new sites. Sensor robustness and thus data reliability are key elements of a monitoring system.

ACKNOWLEDGMENTS

Anna Christler, Theresa Scharl, Dominik G, Sauer, and Michael Melcher have been supported in part by the Federal Ministry for Digital and Economic Affairs (BMWD), the Federal Ministry for Transport, Innovation, and Technology (BMVIT), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, Government of Lower Austria and ZIT–Technology Agency of the City of Vienna through the COMET funding program managed by the Austrian Research Promotion Agency FFG. The funding agencies had no influence on the conduct of this study. The authors thank Manuel Garbe, Anna-Carina Frank, Edit Felföldi and Magdalena Mosor for technical support.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, (A. D.), upon reasonable request.

ORCID

Alois Jungbauer http://orcid.org/0000-0001-8182-7728

Astrid Dürauer http://orcid.org/0000-0002-6007-7697

REFERENCES

Chemalil, L. (2020). Online/at-line measurement, analysis and control of product titer & critical product quality attributes (CQAs) during process development. *Biotechnology and Bioengineering*. *117*, 3757-3765. https://doi.org/10.1002/bit.27531

Christler, A., Felföldi, E., Mosor, M., Sauer, D., Walch, N., Dürauer, A., & Jungbauer, A. (2020). Semi-automation of process analytics reduces operator effect. *Bioprocess and Biosystems Engineering*, *43*(5), 753–764. https://doi.org/10.1007/s00449-019-02254-y

Eilers, P. H., & Boelens, H. F. (2005). Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, *1*(1), 5.

Felfödi, E., Scharl, T., Melcher, M., Dürauer, A., Wright, K., & Jungbauer, A. (2020). Osmolality is a predictor for model-based real time monitoring of concentration in protein chromatography. *Journal of Chemical Technology and Biotechnology*, *95*(4), 1146–1152. https://doi.org/10.1002/jctb.6299

Grote, B., Zense, T., & Hitzmann, B. (2014). 2D-fluorescence and multivariate data analysis for monitoring of sourdough fermentation process. *Food Control*, *38*(1), 8–18. https://doi.org/10.1016/j.foodcont.2013.09.039

Hintersteiner, B., Lingg, N., Janzek, E., Mutschlechner, O., Loibner, H., & Jungbauer, A. (2016). Microheterogeneity of therapeutic monoclonal antibodies is governed by changes in the surface charge of the protein. *Biotechnology Journal*, *11*(12), 1617–1627. https://doi.org/10.1002/biot.201600504

Hintersteiner, B., Lingg, N., Zhang, P., Woen, S., Hoi, K. M., Stranner, S., Wiederkum, S., Mutschlechner, O., Schuster, M., Loibner, H., & Jungbauer, A. (2016). Charge heterogeneity: Basic antibody charge variants with increased binding to Fc receptors. *mAbs*, *8*(8), 1548–1560. https://doi.org/10.1080/19420862.2016.1225642

Jiang, M., Severson, K. A., Love, J. C., Madden, H., Swann, P., Zang, L., & Braatz, R. D. (2017). Opportunities and challenges of real-time release testing in biopharmaceutical manufacturing. *Biotechnology and Bioengineering*, *114*(11), 2445–2456. https://doi.org/10.1002/bit.26383

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3

Liland, K. H., Almoy, T., & Mevik, B.-H. (2010). Optimal choice of baseline correction for multivariate calibration of spectra. *Applied Spectroscopy*, *64*(9), 1007–1016.

Mendhe, R., Thukkaram, M., Patil, N., & Rathore, A. S. (2015). Comparison of PAT based approaches for making real-time pooling decisions for process chromatography: Use of feed forward control. *Journal of Chemical Technology and Biotechnology*, *90*(2), 341–348. https://doi.org/10.1002/jctb.4448

Mevik, B.-H., Wehrens, R., & Liland, K. H. (2019). pls: Partial least squares and principal component regression. *R package version*, *2* 7-2.

Rathore, A. S., Wood, R., Sharma, A., & Dermawan, S. (2008). Case study and application of process analytical technology (PAT) towards bioprocessing: II. Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography. *Biotechnology and Bioengineering*, *101*(6), 1366–1374. https://doi.org/10.1002/bit.21982

Rathore, A. S., Yu, M., Yeboah, S., & Sharma, A. (2008). Case study and application of process analytical technology (PAT) towards bioprocessing: Use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography. *Biotechnology and Bioengineering*, *100*(2), 306–316. https://doi.org/10.1002/bit.21759

Roychoudhury, P., Harvey, L. M., & McNeil, B. (2006). At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy. *Analytica Chimica Acta*, *561*(1–2), 218–224. https://doi.org/10.1016/j.aca.2006.01.037

Sauer, D. G., Melcher, M., Mosor, M., Walch, N., Berkemeyer, M., Scharl-Hirsch, T., Leisch, F., Jungbauer, A., & Dürauer, A. (2019a). Real-time monitoring and model-based prediction of purity and quantity during a chromatographic capture of fibroblast growth factor 2. *Biotechnology and Bioengineering*, *116*(8), 1999–2009. https://doi.org/10.1002/bit.26984

Sauer, D. G., Mosor, M., Frank, A.-C., Weiß, F., Christler, A., Walch, N., Jungbauer, A., & Dürauer, A. (2019b). A two-step process for capture and purification of human basic fibroblast growth factor from E. coli homogenate: Yield versus endotoxin clearance. *Protein Expression and Purification*, *153*, 153–182. https://doi.org/10.1016/j.pep.2018.08.009

Shekhawat, L. K., & Rathore, A. S. (2019). Mechanistic modeling based process analytical technology implementation for pooling in hydrophobic interaction chromatography. *Biotechnology Progress*, *35*(2), e2758. https://doi.org/10.1002/btpr.2758

Team, R. C. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org

Walch, N., Scharl, T., Felföldi, E., Sauer, D. G., Melcher, M., Leisch, F., Dürauer, A., & Jungbauer, A. (2019). Prediction of the quantity and purity of an antibody capture process in real time. *Biotechnology Journal*, *14*(7), 1–10. https://doi.org/10.1002/biot.201800521

Wasalathanthri, D. (2020). Real-time monitoring of quality attributes by in-line fourier transform infrared spectroscopic sensors at ultrafiltration and diafiltration of bioprocess. *Biotechnology and Bioengineering*, *117*, 3766–3774. https://doi.org/10.1002/bit.27532

World Health Organization. (2011). *WHO guidelines on transfer of technology in pharmaceutical manufacturing* (pp. 285–309). WHO. https://www.who.int/medicines/areas/quality_safety/quality_assurance/TransferTechnologyPharmaceuticalManufacturingTRS961Annex7.pdf

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Christler, A., Scharl, T., Sauer, D. G., Köppl, J., Tscheließnig, A., Toy, C., Melcher, M., Jungbauer, A., & Dürauer, A. (2021). Technology transfer of a monitoring system to predict product concentration and purity of biopharmaceuticals in real-time during chromatographic separation. *Biotechnology and Bioengineering*, 118, 3941–3952. https://doi.org/10.1002/bit.27870

Supplementary Material

I) Supplementary figure

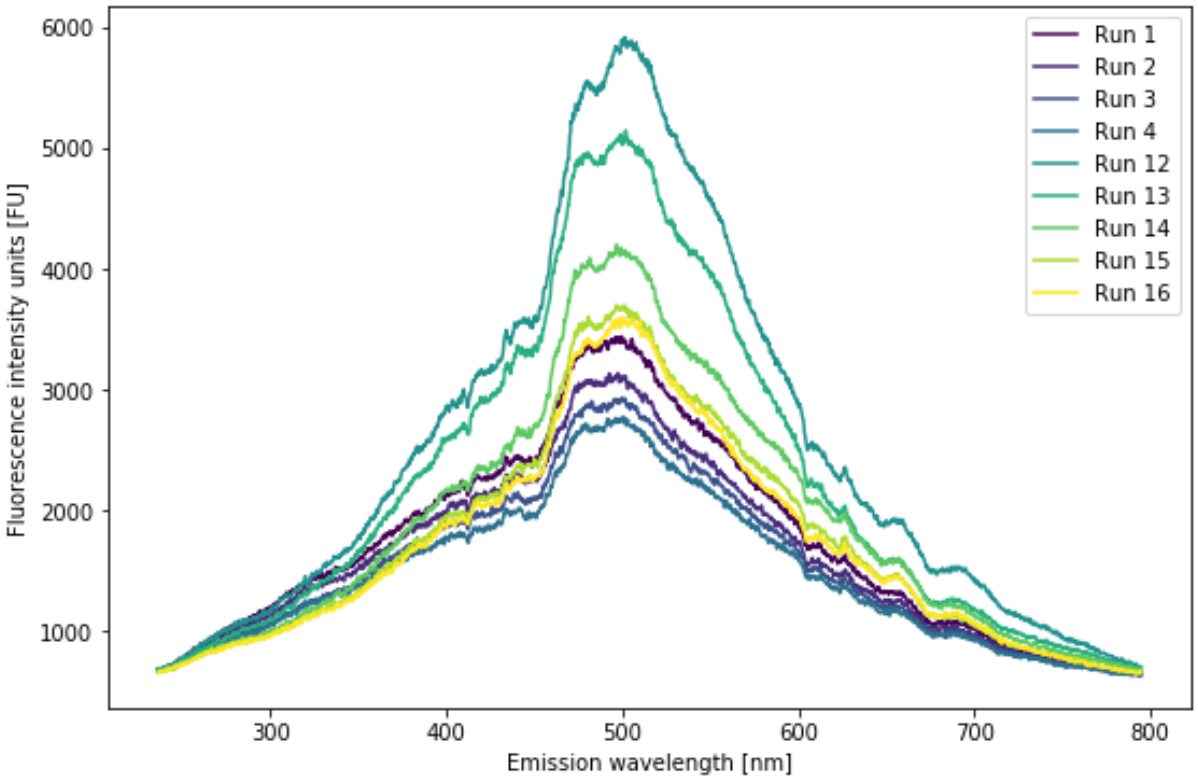


Figure S1: Fluorescence reference signals over a time span of 8 months with fiber exchange after Run 4.

II) Supplementary original data

FGF-2 data (offline measurements and predictions at the three sites) is available from the corresponding author upon request.