Engineering in Life Sciences

WILEY-VCH

Engineering in Life Sciences

RESEARCH ARTICLE OPEN ACCESS

A Novel Raman-Chromatography Assembly for Automated Calibration and In-Line Monitoring in Bioprocessing

Jakob Heyer-Müller[1] | Robin Schiemer[1] | Matthias Lopinski[1] | Caty Wang[1] | Franka Willems[1] | Lars Robbel[2] | Michael Schmitt[2] | Jurgen Hubbuch[1]

[1]Institute of Process Engineering in Life Sciences - Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Baden-Württemberg, Germany | [2]CSL Innovation GmbH Germany, Marburg, Hessen, Germany

Correspondence: Jurgen Hubbuch (juergen.hubbuch@kit.edu)

Received: 25 June 2025 | **Revised:** 11 August 2025 | **Accepted:** 19 August 2025

Keywords: chemometrics | lab automation | Raman spectroscopy | ultrafiltration/diafiltration

ABSTRACT

Optical spectroscopic techniques have been successfully employed in bioprocessing as process analytical technology for real-time process monitoring in numerous applications. The implementation of spectroscopy-based PAT techniques commonly necessitates the generation of representative process data used for calibration and validation of multivariate statistical models for analyzing the sample composition in real-time. To automate the generation of such data, we present a novel assembly of a commercially available chromatography system in combination with a Raman spectrometer for fast and accurate acquisition of Raman spectra. Using the ultra-/diafiltration (UF/DF) process as a case study, our methodology involved the preparation of representative calibration and validation mixtures of phosphate and citrate buffer and lysozyme as a model protein. Chemometric PLS models were calibrated and validated using these datasets, and applied to in-line recorded Raman spectra during a UF/DF experiment. The primary results demonstrated that the novel assembly provides robust and precise offline measurement of Raman spectra, which directly compare with in-line record data. The chemometric PLS models showed good alignment in calibration and validation datasets (R2 and Q2), and could be used to simultaneously monitor the buffer and protein concentrations in real-time during UF/DF. This study provides a simple, commercially available setup for automated acquisition of Raman spectra and demonstrates its straightforward application to bioprocess monitoring.

1 | Introduction

Optical spectroscopic techniques are powerful tools for chemical analysis due to their ability to provide detailed molecular information without the need for extensive sample preparation. Optical spectroscopy is often applied in combination with multivariate statistical modeling as a process analytical technology (PAT) for analyzing sample composition in chemical or pharmaceutical processes. In biopharmaceutical downstream processing, the application of Raman spectroscopy has been successfully applied, for example, for the analysis of chromatography [1], filtration [2], or precipitation [3] processes.

The implementation of spectroscopy-based PAT tools into downstream processes requires the generation of representative data for training, optimization, and testing of statistical chemometric models, which is time-consuming and labor-intensive. Representative datasets may be generated by systematic variation

Jakob Heyer-Müller and Robin Schiemer contributed equally to this work.

of the operating conditions of the process of interest [3, 4], artificial spiking with the target analytes or synthetic mixtures [5–7]. The use of synthetic mixtures, when available, facilitates an offline model calibration [8] or can be used to supplement process-derived datasets [5].

To streamline offline model calibration and improve reproducibility, automated sample preparation and data collection are essential. While ultraviolet/visible (UV/Vis) spectra can easily be acquired in a high-throughput mode using plate readers [9], vibrational spectroscopy data such as Raman spectra strongly depend on acquisition settings and the experimental setup. If the chemometric model is intended for online or in-line use, offline model calibration using the same experimental setup is desirable to avoid error-prone data correction methodologies [10]. Recent advancements in automation and data processing have shown promise in addressing these limitations. For example, a custom integrated Raman–Tecan system was developed to automate the calibration process, generating Raman-based models suitable for in-line usage [11].

Monitoring biopharmaceutical ultra-/diafiltration (UF/DF) processes through analytical tools in order to quantify or control the Donnan effect has long been a goal in industry and academia [8, 12, 13]. Monitoring the solution composition in UF/DF processes, including protein and excipient concentrations, has been realized through combinations of multiple sensors [14], near-infrared (NIR) spectroscopy [8], or more recently Raman spectroscopy [2]. Thakur et al. [8] have shown the suitability of offline calibrated chemometric models to monitor UF/DF processes in real-time. Rolinger et al. [2] combined Raman-based chemometric models with a simple mass-balance model through a Kalman filtering approach to increase accuracy of their model's predictions.

Our study introduces a novel assembly of a commercially available ÄKTA chromatography system in combination with an autosampler and a Raman spectrometer to automate Raman measurements with a minimal sample volume of 400 µL. A custom software is used to control the ÄKTA chromatography system via its open platform communication—unified architecture (OPC-UA) interface. Two datasets involving synthetic mixtures of phosphate- and citrate-buffered lysozyme solutions are prepared using different statistical designs. Partial least square (PLS) regression models are calibrated using the data from the synthetic mixtures. The calibrated PLS models are then applied to in-line Raman spectra from UF/DF involving the same buffer–protein matrix. The goals of this study are to demonstrate the feasibility and effectiveness of automating Raman spectroscopy measurements using a chromatography system for straightforward model calibration. The article outlines the process from system integration and sample preparation to model calibration and validation, showcasing the potential of this automated approach to enhance the efficiency and accuracy of Raman spectroscopy measurements. Our results indicate that the offline calibrated, chemometric PLS models showed good in-line performance, following expected trends of the UF/DF experiment, demonstrating the capability of the automated system and the developed models to accurately monitor and predict the concentration of analytes in real-time.

2 | Materials and Methods

2.1 | Experimental Setup

An ÄKTApurifier system equipped with an autosampler A-905 was attached to a Raman spectrometer to automatically inject samples into the flow cell of the Raman spectrometer. A schematic depiction of this setup is shown in Figure 1A. The system was controlled using a custom-made software communicating via OPC-UA with the ÄKTA chromatography system. A schematic communication protocol is depicted in Figure 1B. To enable automated equilibration, loading, injection, measurement, and washing, the valves and tubing connections were chosen as visualized in Figure 1C–G. For all injections in this study, an injection volume of 490 µL was used. The protocol for a single injection involved 8 mL of system equilibration with ultrapure water (PURELAB Ultra, ELGA LabWater), followed by the injection of the sample solution into the Raman spectrometer flow cell. While the measurement was ongoing, the flow rate of the system was set to 0 mL/min. This was done to minimize the volume needed per injection as measurements under flow would require a multiple of the applied 400 µL. Each measurement was performed for 29 s totaling to an acquisition of 24 Raman spectra (see Section 2.2 for more information on the Raman spectroscopy measurements). Afterward, the sample was washed out using 0.1 M NaOH for 2 mL and a subsequent re-equilibration with ultrapure water for 3 mL. All phases were performed using a flow rate of 2 mL/min.

2.2 | Raman Spectroscopy

A Hyperflux PRO Plus (Tornado Spectral Systems, Toronto, Canada) equipped with a 785 nm laser, a Bioreactor BallProbe, and a 200 µL flow cell (both MarqMetrix, Seattle, USA) were used. Raman spectra were recorded at a laser power of 495 mW and an exposure time of 1200 ms. For the offline data, from the 24 exposures the first and last two were excluded, leading to a total of 20 exposures per spectrum. For the in-line data, the same setting as for the offline data was used, but only 10 exposures were averaged per spectrum.

2.3 | Case Study: Sample Preparation

Samples of 600 µL were automatically prepared in 2 mL deep well plates using a Freedom Evo liquid handling station (Tecan, Männedorf, Switzerland), containing different concentrations of citrate and phosphate buffer and lysozyme. Two datasets were prepared for calibration and validation of chemometric models using different experimental designs, namely, a Latin hypercube sampling (LHS) and a central composite design (CCD). The LHS and CCD sets were composed of 32 and 18 conditions with 3 replicates per condition, totaling at 96 and 54 samples, respectively. The compositions of all samples of the LHS and CCD datasets are illustrated in Figure 4. All samples were prepared using stock solutions of 0.3 M sodium phosphate buffer (pH 7.0), 0.4 M sodium citrate buffer (pH 5.8), and 50 mg/mL lysozyme dissolved in 0.1 M sodium chloride. All solutions were prepared with ultrapure water (PURELAB Ultra, ELGA LabWater).

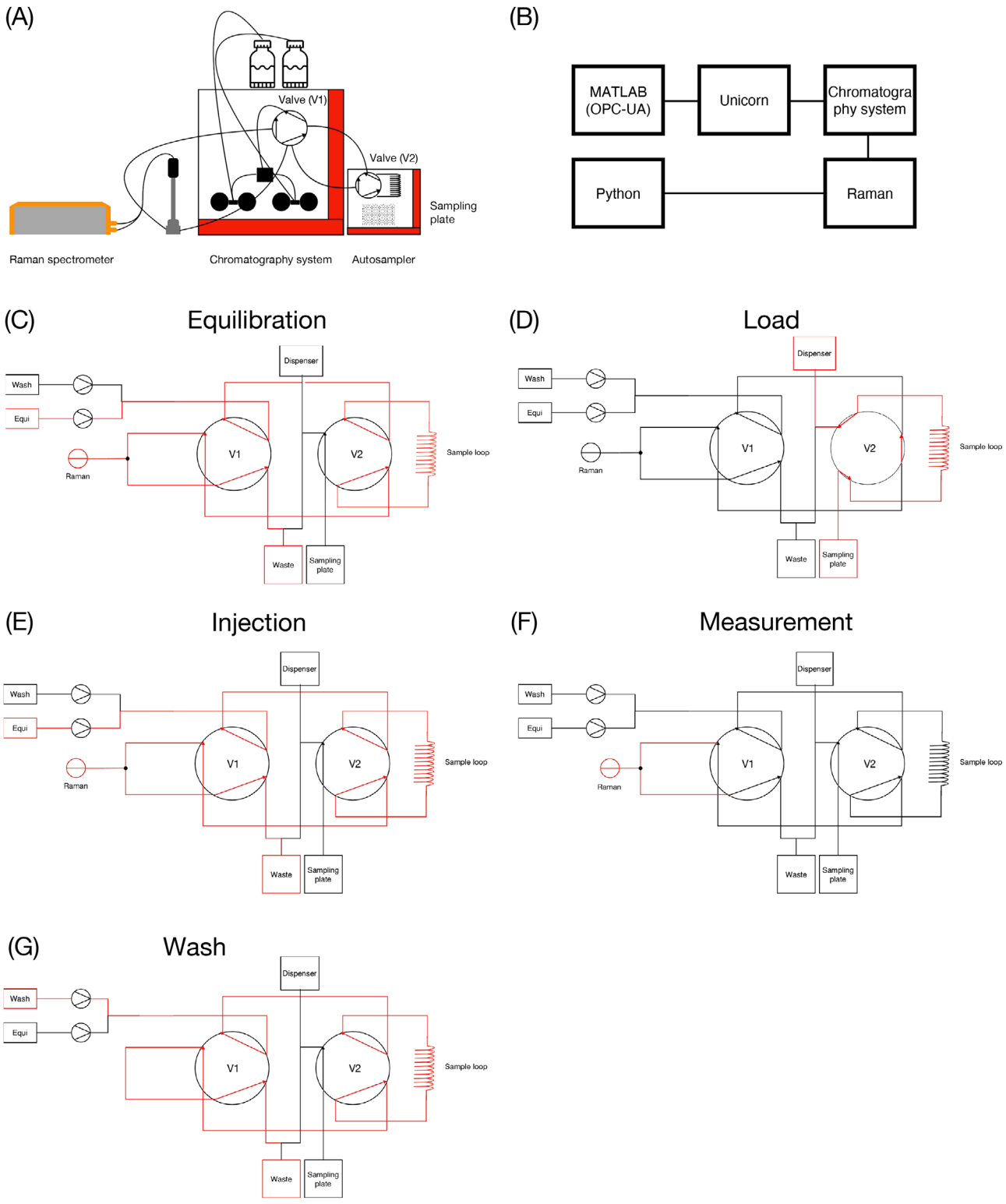

FIGURE 1 | Experimental setup connecting ÄKTA purifier, A905 autosampler, and Tornado Hyperflux PRO Plus. Schematic illustration of the experimental setup (A). Schematic depiction of software interfaces (B). Flow paths for different phases: equilibration (C), sample load (D), sample injection (E), measurement (F) and wash (G). Red and black lines indicate active and non-active flow paths, respectively.

2.4 | Data Preprocessing and Chemometric Modeling

Raman spectra were preprocessed by a sequence of operations. First, all spectra were cropped to wavenumbers from 500 to 1800 cm^{-1}. The spectra were then smoothed using a Savitzky-Golay (SG) filter with a window size of 21 and polynomial order of 3. All spectra were subsequently normalized using the sapphire band at 750 cm^{-1}. Due to discrepancies in absolute intensity between offline and in-line data, which most likely occurred

due to differences in the lysozyme batch, a direct standardization protocol was used according to Nikzad-Langerodi and Andries [15]. As no direct calibration samples were available, synthetic spectra representing the starting condition of the in-line experiments were generated using the local subset augmentation (LSA) algorithm presented in Schiemer et al. [16]. The direct standardization problem was formulated as

$$X_{\mathrm{T}}W - X_{\mathrm{S}} + \lambda_{\mathrm{reg}} * D^{\mathrm{T}}\mathrm{D} \quad (1)$$

where $\mathbf{X}_{\mathrm{T}}$ and $\mathbf{X}_{\mathrm{S}}$ are the target and source spectra, respectively. The equation is further regularized using the second-order difference matrix $\mathbf{D}$ and a regularization factor $\lambda_{\mathrm{reg}} = 0.001$. The residual between the target and source spectra is minimized by determining the weight matrix $\mathbf{W}$. The direct standardization procedure is applied for each experimental run individually. The weight matrix $\mathbf{W}$ determined by solving the direct standardization problem for the first spectrum was then used to transform the remaining in-line spectra via $\mathbf{X'}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\mathbf{W}$ of the corresponding experiment.

Finally, baseline drifts were corrected in both the offline and in-line data using a derivative peak-screening asymmetric least squares algorithm (DERPSALSA)-based Whittaker filter as implemented in the pybaselines library (v1.1.0) and proposed in Korepanov [17]. The Whittaker filter was configured to $\lambda = 10^6$, $k = 0.05$, and $p = 0.01$. The parameters were determined by visual comparison and yielded comparable results for both the in-line and offline data. An illustrative overview of the preprocessing and standardization procedure can be found in Figure S1. The article further uses normalized intensity in all displays of spectral data. The normalized intensity refers to the values obtained after the entire sequence of preprocessing operations and should not be confused with the typical notion of normalized values ranging from 0 to 1.

Chemometric models were calibrated using the LHS dataset, and their performance was evaluated using the CCD sets. All PLS models were used as implemented in scikit-learn (v1.5.2). The number of latent variables was optimized by K-Fold cross-validation of the LHS data using 3 folds. The optimal number of latent variables was chosen based on the scaled RMSECV according to Wold et al. [18]. Before being fed into the PLS models, all preprocessed spectra were treated by a first-derivative SG filter with a window size of 11 and a second-order polynomial and subsequently mean-centered. For the lysozyme model, the spectra were finally cropped to the wavenumber range from 900 to 1300 cm^{-1} to reduce the co-influence of the other species. Separate PLS models were used for each target species, namely, phosphate, citrate, and lysozyme. All computations were done in Python 3.10.

2.5 | Ultrafiltration/Diafiltration

To evaluate the calibrated PLS models for in-line use, two UF/DF experiments were conducted at starting concentrations of approx. 10 and 5 g/L lysozyme. To further increase variability, the starting concentration of phosphate and the citrate concentration in the DF buffer were varied alongside the protein concentration. A summary of the process conditions is presented in Table 1. For UF/DF, a KrosFlo KRIIi (Spectrum Labs) cross-flow filtration (CFF) system was used. A MicroKros Hollow Fiber filter module with a molecular weight cutoff (MWCO) of 3 kDa and a membrane area of 20 cm^2 was used. The transmembrane pressure (TMP) was controlled using an automated valve in the retentate stream. The retentate flow was set to 45 mL/min. The Raman spectrometer was connected via a separate online loop with an Minipuls 3 peristaltic pump (Gilson, Villiers le Bel, France). The flow rate in the online loop was set to 1 mL/min and monitored using a SLS-1500 flow meter (Sensirion, Stafa, Switzerland). The online loop circulated the process solution from and back into the retentate reservoir container. A detailed scheme of the experimental setup is shown in Figure 2. During each experiment, the starting solution was prepared and circulated in the CFF system before applying TMP. For the protein runs, the first ultrafiltration (UF) phase was conducted until a concentration factor of approx. 2 has been reached. The DF phase was initiated by manually attaching the DF buffer. After approx. 3 diafiltration volumes (DVs), the second UF phase was initiated. The process was terminated when 10 g permeate mass had been collected after start of the second UF phase.

TABLE 1 | Experimental conditions for different runs.

Experiment	Lysozyme (g/L)	Phosphate (M)	Citrate (M)	Pressure (bar)
Run 1	11.5	37.5	25	0.4
Run 2	5.7	35.0	75	0.8

3 | Results and Discussion

3.1 | Automating Raman Spectroscopy Measurements by Combining an ÄKTA Chromatography System With a Raman Spectrometer

To characterize the experimental setup for Raman measurements and evaluate robustness of the system, multiple injection sequences were performed. In Figure 3A, the Raman measurements of a lysozyme dilution series are exemplarily depicted. The sequence shows multiple subsequent injections at different lysozyme concentrations with intermittent washing cycles to re-equilibrate the flow cell. The depicted signal at 1004 cm^{-1} was extracted from Raman spectra after preprocessing according to Section 2.4. After each injection the signal reduces to a stable baseline, supporting the assumption of a sufficiently long washing phase and complete re-equilibration. In Figure 3B, a single injection is depicted in detail. The sample is injected into the flow cell using the autosampler. The vertical lines indicate the time interval during which the Raman spectra are collected for the injected sample. During this phase, the flow is halted (0 mL/min) for the collection of the Raman spectra and is thus considered a static measurement. Halting the flow is a strict necessity to keep the injection volume minimal. With the current setup, dynamic measurements would only be possible by applying a flow from the system pump, which increases the necessary injection volume. By integration of an external

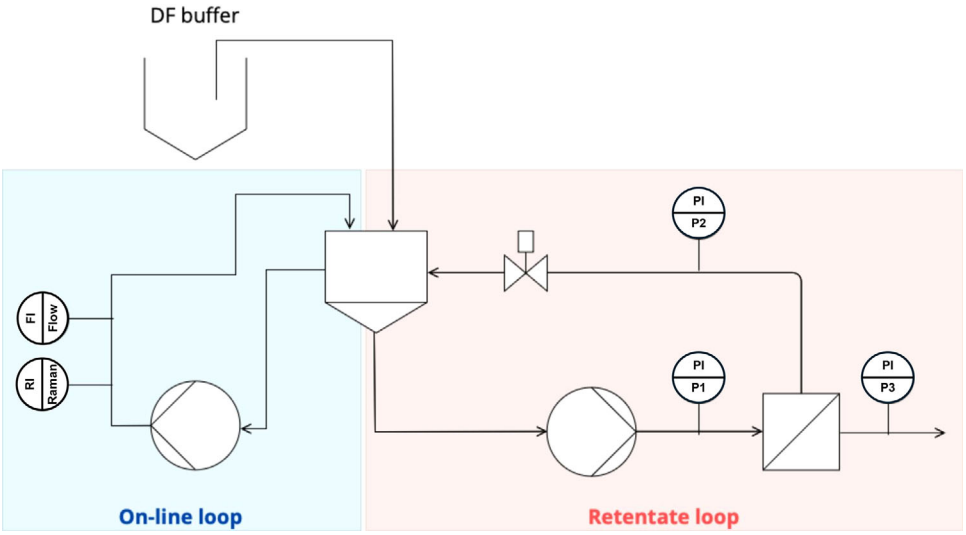

FIGURE 2 | Process flow diagram of the crossflow filtration (CFF) system with an external loop for online Raman measurements. DF buffer: diafiltration buffer. P1: Feed pressure. P2: Retentate pressure. P3: permeate pressure. Flow: flow meter. Raman: Raman spectrometer.

pump in the measurement loop, dynamic measurements would be possible. During the injection and washout of the samples, a continuous flow is applied to the analyte causing a small overshoot in the averaged Raman signal. This overshoot is mainly caused by an amplified baseline in the Raman spectra. Noticeably, the 1004 cm^{-1} band after baseline correction is not affected by this phenomenon. Within the measurement interval, the Raman signal intensity is scattering due to the intrinsic variability of the Raman spectrometer in combination with the algorithmic preprocessing. This variability is known to originate from the detector and shot noise [19], which influence both the total intensity and the preprocessing performance. The static measurement phase was configured to last 24 exposures for each sample measured in this study, to ensure stable signals after averaging. In Figure 3C, an overlay of Raman spectra for dilution series of sodium phosphate (pH 7), sodium citrate (pH 5.8), and lysozyme is depicted in shades of blue, green, and red, respectively. The spectra are shown after preprocessing. In the spectral region between 600 and 1800 cm^{-1}, all analytes show specific Raman profiles with strongly overlapping peaks. While phosphate and citrate dominate the Raman spectra at Raman shifts of 989 and 1417 cm^{-1}, the Raman bands of lysozyme are comparably smaller. The lysozyme band at 1350 cm^{-1} is the least interfered by the other species in the mixture. This high degree of overlap underlines the need for multivariate regression models, such as PLS models, as univariate models are not capable to track the analyte concentrations under varying sample composition. The proposed setup provides an easy way to repurpose chromatography systems for automated spectral acquisition. While other laboratory systems, like high-performance liquid chromatography (HPLC) systems, come with similar built-in functionality, the flow cell of the Raman spectrometer was built for preparative lab-scale experiments and hence is not compatible with the tubing dimensions of an HPLC system. Furthermore, the setup allows for measurement with other spectrometers or incorporation of an external pump to allow for dynamic measurements, which would increase representativeness of the acquired data with in-process data.

3.2 | Calibrating Chemometric Models Off Line by Synthetic Mixtures

After successful characterization, the presented assembly was used to record Raman data for the monitoring of a UF/DF process using lysozyme as a model protein. To build PLS models for process monitoring, the data used for training and testing require the coverage of the process design space while ensuring statistical independence of all target species [20]. In this case, the target species were sodium phosphate, sodium citrate, and lysozyme. For calibration and validation of the PLS models, the LHS and CCD datasets were used, respectively. Figure 4 presents the sample composition, correlation matrices of the target species, and preprocessed Raman spectra for the LHS and CCD datasets. The LHS and CCD dataset contains 32 and 18 different sample compositions, respectively, with three replicates each. Although the LHS dataset, as a randomized space-filling design, covers the process design space with minimal target correlation (R^2 = 0.2), the CCD dataset perfectly isolates the effects of the target species (R^2 = 0). The LHS and CCD designs were chosen to be able to calibrate the PLS models in the desired design space and independently validate the PLS models using a separate dataset of a different statistical scheme. The corresponding Raman spectra in Figure 4C–F are colored from low to high lysozyme concentrations in purple to yellow, respectively. The spectra show the covariation introduced by varying the buffer concentrations and corresponding pH. As previously described in Section 3.1, the three analytes largely overlap across the spectral range with the region around 1200–1400 cm^{-1} showing the least overlap of the lysozyme and the buffer species. Figure 5 presents parity plots for training and test data for all three target species. In all cases, the PLS models show good alignment between theoretical and predicted values for training and test data. The lysozyme model was trained in range of 0 to 20 g/L with an R^2 of 0.99. The accuracy for the test set reduces with a Q^2 of 0.87 and an RMSEP of 1.88 corresponding to 2.5-fold increase compared to the training data. This is mainly due to a consistent overestimation of the lysozyme concentration for samples with

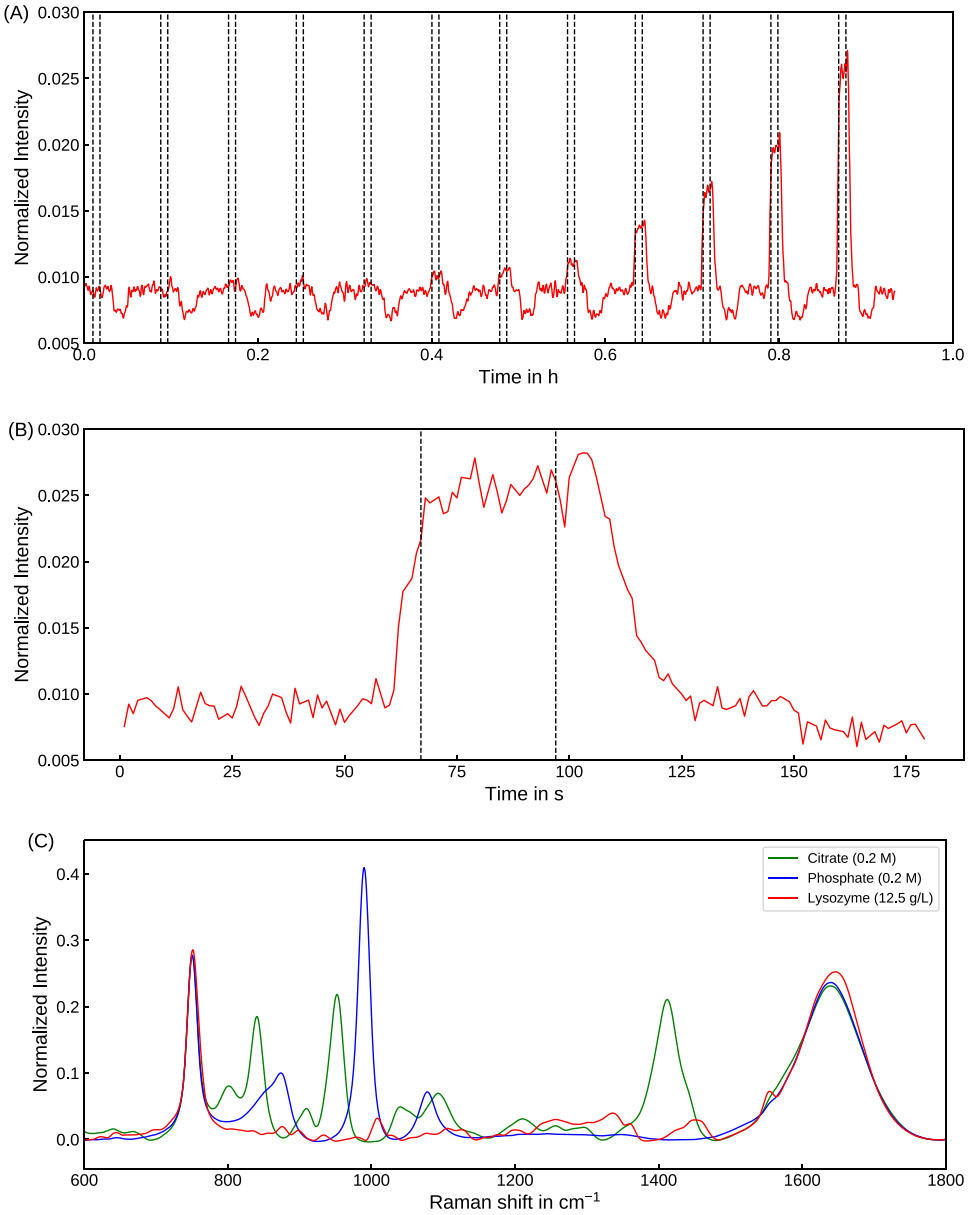

FIGURE 3 | Injection sequence of selected samples with increasing lysozyme concentration from 0.175 to 12.5 g/L depicted by the Raman intensity using the 1004 cm^{-1} band after baseline correction as an indicator (A). The Raman intensity was normalized to the sapphire band at 750 cm^{-1} prior to baseline correction. Dashed vertical lines indicate the time intervals for the static measurements. Zoomed Raman intensity profile for the single injection of the 12.5 g/L sample using the 1004 cm^{-1} band after baseline correction and the same normalization approach (B). Baseline-corrected Raman spectra for samples of 12.5 g/L lysozyme (red), 0.2 M sodium phosphate pH 7.0 (blue), and 0.2 M sodium citrate pH 5.6 (green) (C) using the same normalization approach.

concentrations above 10 g/L. The reason for this offset is most likely a discrepancy between the training and test set itself as they were prepared using two different lysozyme stock solutions, suggesting that for robust model calibration it would be advisable to repeat the procedure multiple times to include potential variations. Often, chemometric models are validated on random splits of the same dataset [8], which may lead to too optimistic estimation of model errors and hence may negatively affect model performance when used in their intended application. We thus always recommend to record multiple independent datasets to test chemometric models. In this scenario, despite the overestimation of the lysozyme concentrations, the CCD dataset additionally evaluates the influence of the other target species that are varied independently from the lysozyme concentration. The phosphate and citrate model predictions show almost perfect alignment with the theoretical concentrations with RMSEC 1.27 and 1.75 mM, respectively. In contrast, to the lysozyme model, the test errors only increased approx. 2-fold compared to the training errors to RMSEP of 2.93 mM for both species. In comparison to Thakur et al. [8], we achieve comparable accuracy as has been demonstrated for other common buffer systems like histidine and acetate. For quantification of monoclonal antibody (mAb) concentration, the authors state an RMSEP of 3.25 g/L, whereas their model was calibrated over a wider design space, spanning 0 to 200 g/L. Noticeably, the changing pH in the range of 5.8 to 7.0 in the analyzed sample within the LHS and CCD datasets did not

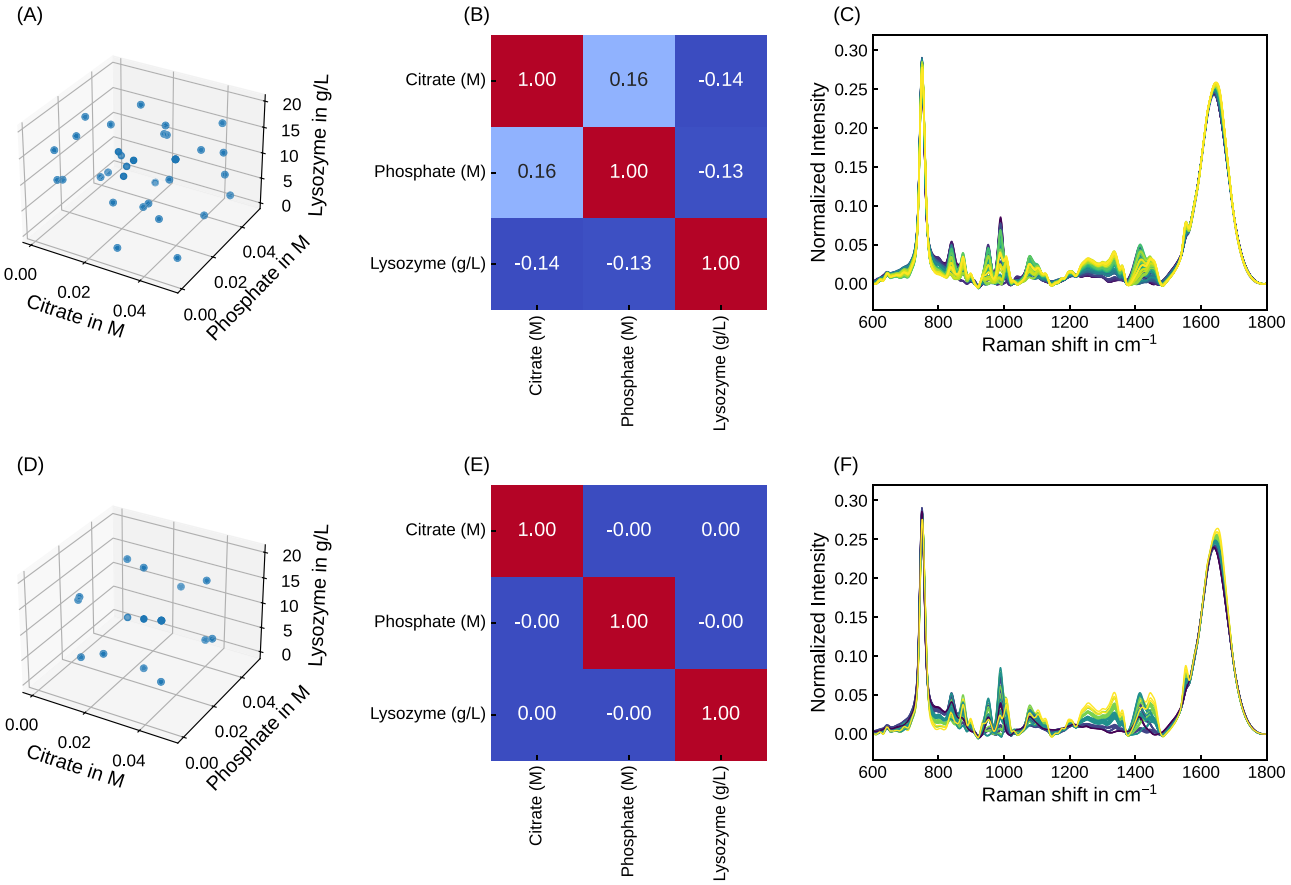

FIGURE 4 | Sample distribution (A, D), correlation coefficients (B, E), and preprocessed Raman spectra (C), (F) for LHS and CCD datasets in the upper and lower row, respectively.

negatively affect the model performance, even though pH affects the Raman spectra of buffer substances(21.

3.3 | Applying Chemometric Models to Quantify Protein and Buffer Concentrations in Ultrafiltration/Diafiltration Processes

The established PLS models were applied to Raman data, which were recorded online through an external loop sourcing from the UF/DF reservoir as depicted in Figure 2. Data acquisition was started once the reservoir and retentate loop were filled with the load material. In total, two runs at starting concentrations of 11.5 and 5.7 g/L were performed. The detailed experimental conditions are listed in Table 1. Figure 6 presents model predictions for Runs 1 and 2 in (A) and (B), respectively. It needs to be noted that for these experiments, no offline reference measurements were available to quantify the actual content of the substances. However, from the predicted trajectories a qualitative evaluation can be done. The UF/DF runs 1 and 2 can be broadly separated into the UF1, DF, and UF2 phase. During UF1 and UF2, the concentrations of phosphate and citrate are expected to remain roughly constant while the concentration of lysozyme is expected to increase. During the DF phase, the lysozyme concentration is expected to remain constant while a buffer exchange from phosphate to citrate is performed. The three phases and the described process dynamics are well represented in the Raman model predictions. In both runs, the PLS models predict an increase in the lysozyme concentration from initially 11 and 5 g/L to approx. 25 and 15 g/L after UF1 and 30 and 20 g/L after UF2, respectively. Predictions for phosphate and citrate remain mostly constant during UF1 and UF2. During the DF phase, the phosphate concentration quickly decreases from initially 35 and 40 mM until a complete depletion is reached. Inversely, the citrate concentration is predicted to increase and reaches a plateau of 35 and 90 mM at the end of the DF phase in runs 1 and 2, respectively. The lysozyme concentration remains mostly stable over the DF phase with minor variations such as an overall decrease in run 1 and an initial drop in run 2. These effects may be caused by product sieving through the UF/DF membrane or the sensitivity of the model to other species. Although the PLS models produce realistic predictions of the investigated analyte concentrations, several limitations can be pointed out. Due to the discrepancy between the offline and online collected data, which could not be entirely eliminated by the direct standardization approach, the phosphate model predicts negative concentrations, and the citrate model likely overestimates the citrate content at the end of the DF phase. This suggests that for successful implementation of Raman spectroscopy for UF/DF monitoring, online data should be incorporated into model training. Furthermore, the chosen experimental conditions do not match realistic operating conditions in biopharmaceutical UF/DF processes, where concentrations of up to 200 g/L are typical for formulated mAb solutions [22]. Under these conditions, the Donnan effect causes macroscopic pH shifts and retention of buffer substances [23]. These effects were out of scope for this study, and the capability of Raman spectroscopy to

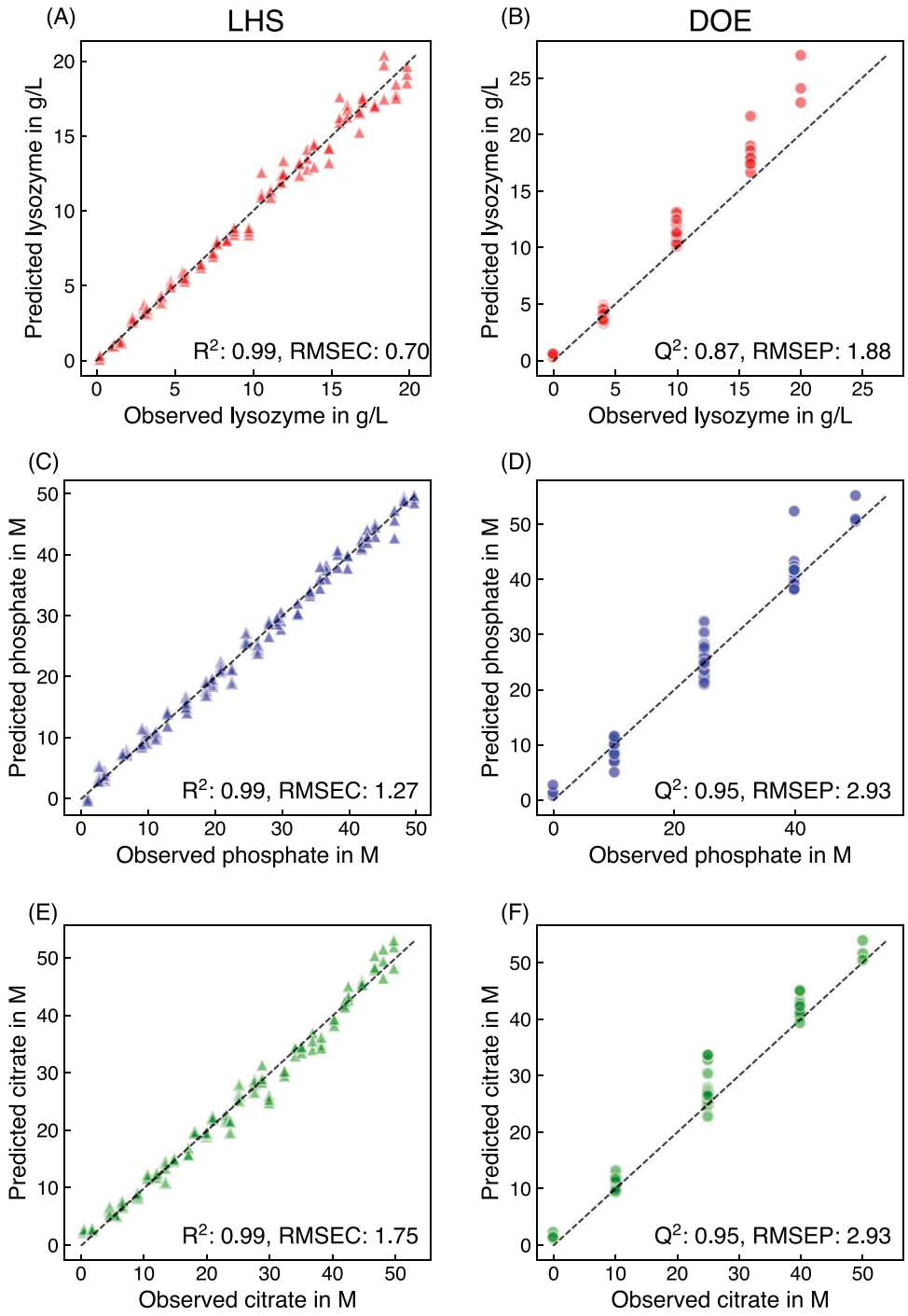

FIGURE 5 | Regression errors for calibration datasets LHS and CCD. The errors are displayed for lysozyme, phosphate, and citrate from top to bottom. R^2 and RMSEC values refer to the error metrics obtained from the data assigned to training, while Q^2 and RMSEP refer to the data assigned to test.

monitor buffer concentrations influenced by the Donnan effect will be evaluated in future studies.

4 | Conclusions

This study demonstrates the feasibility of automating Raman spectroscopy measurements using a chromatography system with a minimal sample volume of 490 µL. A custom software interface enabled seamless integration of the ÄKTA chromatography system with a Raman spectrometer, allowing for efficient data acquisition and chemometric model calibration. The developed system was evaluated using synthetic mixtures of phosphate- and citrate-buffered lysozyme solutions, with calibration performed through PLS regression models. The PLS models trained on offline data demonstrated good alignment with theoretical values, with minor deviations in high-concentration lysozyme samples. When applied in-line to UF/DF processes, the models accurately

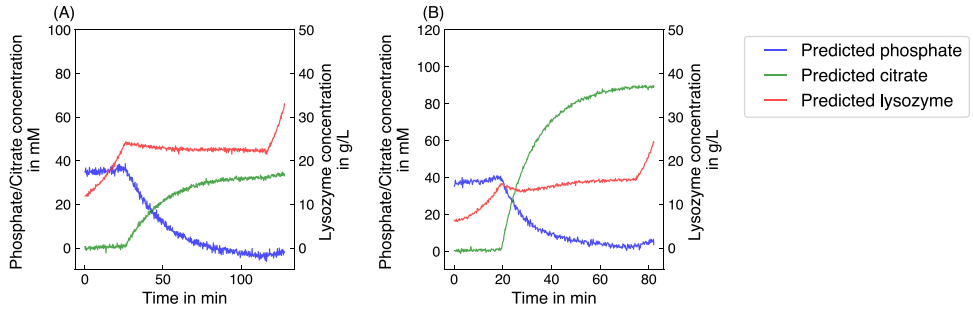

FIGURE 6 | Model predictions for in-line data from UF/DF experiments.

reflected expected concentration trends of lysozyme and buffer components. However, discrepancies between offline and online data led to limitations in phosphate and citrate predictions, highlighting the need for online data incorporation in future model calibration. Despite these limitations, the system successfully captured the key dynamics of UF/DF processes, demonstrating its potential for real-time monitoring of protein and buffer concentrations. Further research should address the impact of process conditions, such as high protein concentrations and Donnan effects, to enhance model robustness and applicability in biopharmaceutical settings.

Author Contributions

Jakob Heyer-Müller and Robin Schiemer both evolved the concepts and methods presented in this article, supervised experimental work, analyzed and interpreted the data, drafted the figures, and wrote the final manuscript. Matthias Lopinski developed automation software and performed initial experiments. Franka Willems and Caty Wang both conducted experiments for UF/DF case study. Lars Robbel and Michael Schmitt, and Jurgen Hubbuch supervised the work. Jürgen Hubbuch acquired funding and initiated the work. All authors read and approved the final manuscript.

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting this study's findings, as well as the OPC-UA-based communication software, are available from the corresponding author upon reasonable request.

References

1. F. Feidl , S. Garbellini, M. F. Luna, et al., "Combining Mechanistic Modeling and Raman Spectroscopy for Monitoring Antibody Chromatographic Purification," *Processes* 7, no. 10 (2019): 683.

2. L. Rolinger, J. Hubbuch, and M. Rudt, "Monitoring of Ultra- and Diafiltration Processes by Kalman-Filtered Raman Measurements," *Analytical and Bioanalytical Chemistry* 415, no. 5 (2023): 841–854.

3. A. Dietrich, R. Schiemer, J. Kurmann, S. Zhang, and J. Hubbuch, "Raman-Based PAT for VLP Precipitation: Systematic Data Diversification and Preprocessing Pipeline Identification," *Frontiers in Bioengineering and Biotechnology* 12 (2024): 1399938.

4. N. Brestrich, M. Rudt, D. Büchler, and J. Hubbuch, "Selective Protein Quantification for Preparative Chromatography Using Variable Pathlength UV/Vis Spectroscopy and Partial Least Squares Regression," *Chemical Engineering Science* 176 (2018): 157–164.

5. M. Klaverdijk, M. Ottens, and M. E. Klijn, "Single Compound Data Supplementation to Enhance Transferability of Fermentation Specific Raman Spectroscopy Models," *Analytical and Bioanalytical Chemistry* 417 (2025): 1873–1884.

6. P. Romann, J. Kolar, D. Tobler, C. Herwig, J. Bielser, and T. K. Villiger, "Advancing Raman Model Calibration for Perfusion Bioprocesses Using Spiked Harvest Libraries," *Biotechnology Journal* 17, no. 11 (2022): e2200184.

7. R. M. Santos, J. Kessler, P. Salou, J. C. Menezes, and A. Peinado, "Monitoring mAb Cultivations With In-Situ Raman Spectroscopy: The Influence of Spectral Selectivity on Calibration Models and Industrial Use as Reliable PAT Tool," *Biotechnology Progress* 34, no. 3 (2018): 659–670.

8. G. Thakur, V. Hebbi, and A. S. Rathore, "Near Infrared Spectroscopy as a PAT Tool for Monitoring and Control of Protein and Excipient Concentration in Ultrafiltration of Highly Concentrated Antibody Formulations," *International Journal of Pharmaceutics* 600 (2021): 120456.

9. S. K. Hansen, B. Jamali, and J. Hubbuch, "Selective High Throughput Protein Quantification Based on UV Absorption Spectra," *Biotechnology and Bioengineering* 110, no. 2 (2013): 448–460.

10. J. Wang, J. Chen, J. Studts, and G. Wang, "Automated Calibration and In-Line Measurement of Product Quality During Therapeutic Monoclonal Antibody Purification Using Raman Spectroscopy," *Biotechnology and Bioengineering* 120 (2023): 3288–3298.

11. J. Chen, J. Wang, R. Hess, G. Wang, J. Studts, and M. Franzreb, "Application of Raman Spectroscopy During Pharmaceutical Process Development for Determination of Critical Quality Attributes in Protein A Chromatography," *Journal of Chromatography A* 1718 (2024): 464721.

12. J. Abel, A. Kosky, N. Ball, H. Bacon, R. Kaushik, and G. R. Kleemann, "A Small-Scale Process for Predicting Donnan and Volume Exclusion Effects During Ultrafiltration/Diafiltration Process Development," *Journal of Pharmaceutical Sciences* 107, no. 5 (2018): 1296–1303.

13. J. M. West, H. Feroz, X. Xu, et al., "Process Analytical Technology for On-Line Monitoring of Quality Attributes During Single-Use Ultrafiltration/Diafiltration," *Biotechnology and Bioengineering* 118, no. 6 (2021): 2293–2300.

14. M. Rüdt, L. Rolinger, J. Diehm, J. Chow-Hubbertz, et al., "Multi-Attribute PAT for UF/DF of Proteins—Monitoring Concentration, Particle Sizes, and Buffer Exchange," *Analytical and Bioanalytical Chemistry* 412, no. 9 (2020): 2123–2136.

15. R. Nikzad-Langerodi and E. Andries, "A Chemometrician's Guide to Transfer Learning," *Journal of Chemometrics* 35, no. 11 (2021): e3373.

16. R. Schiemer, M. Rüdt, and J. Hubbuch, "Generative Data Augmentation and Automated Optimization of Convolutional Neural Networks

for Process Monitoring," *Frontiers in Bioengineering and Biotechnology* 12 (2024): 1228846.

17. V. I. Korepanov, "Asymmetric Least-Squares Baseline Algorithm With Peak Screening for Automatic Processing of the Raman Spectra," *Journal of Raman Spectroscopy* 51, no. 10 (2020): 2061–2065.

18. S. Wold, M. Sjöström, and L. Eriksson, "PLS-Regression: A Basic Tool of Chemometrics," *Chemometrics and Intelligent Laboratory Systems* 58, no. 2 (2001): 109–130.

19. J. M. Smulko, N. C. Dingari, J. S. Soares, and I. Barman, "Anatomy of Noise in Quantitative Biological Raman Spectroscopy," *Bioanalysis* 6, no. 3 (2014): 411–421.

20. L. Rolinger, M. Rudt, and J. Hubbuch, "A Critical Review of Recent Trends, and a Future Perspective of Optical Spectroscopy as PAT in Biopharmaceutical Downstream Processing," *Analytical and Bioanalytical Chemistry* 412, no. 9 (2020): 2047–2064.

21. D. H. Müller, M. Börger, J. Thien, and H.-J. Koß, "The Good pH Probe: Non-Invasive pH In-Line Monitoring Using Good Buffers and Raman Spectroscopy," *Analytical and Bioanalytical Chemistry* 415, no. 29 (2023): 7247–7258.

22. F. Rischawy, T. Briskot, F. Nitsch, et al., "Modeling of Biopharmaceutical UF/DF From Laboratory to Manufacturing Scale," *Computers & Chemical Engineering* 177 (2023): 108337.

23. T. Briskot, N. Hillebrandt, S. Kluters, et al., "Modeling the Gibbs–Donnan Effect During Ultrafiltration and Diafiltration Processes Using the Poisson–Boltzmann Theory in Combination With a Basic Stern Model," *Journal of Membrane Science* 648 (2022): 120333.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: elsc70044-sup-0001-SupMat.pdf **Supporting File 2**: elsc70044-sup-0002-FigureS1.pdf