Received: 20 April 2022 | Revised: 14 July 2022 | Accepted: 26 July 2022

DOI: 10.1002/biot.202200184

Biotechnology Journal

RESEARCH ARTICLE

Advancing Raman model calibration for perfusion bioprocesses using spiked harvest libraries

Patrick Romann[1,2] | **Jakub Kolar**[1,3] | **Daniela Tobler**[1] | **Christoph Herwig**[2] | **Jean-Marc Bielser**[4] | **Thomas K. Villiger**[1]

[1]Institute for Pharma Technology, School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland

[2]Research Division Biochemical Engineering, Institute of Chemical Environmental and Bioscience Engineering, Vienna University of Technology, Vienna, Austria

[3]University of Chemistry and Technology Prague, Prague, Czechia

[4]Biotech Process Sciences, Merck Serono SA (an affiliate of Merck KGaA, Darmstadt, Germany), Corsier-sur-Vevey, Switzerland

Correspondence
Thomas K. Villiger, Institute for Pharma Technology, School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland, 4132 Muttenz, Switzerland
Email: thomas.villiger@fhnw.ch

Abstract

Background: Raman spectroscopy has gained popularity to monitor multiple process indicators simultaneously in biopharmaceutical processes. However, robust and specific model calibration remains a challenge due to insufficient analyte variability to train the models and high cross-correlation of various media components and artifacts throughout the process.

Main Methods: A systematic Raman calibration workflow for perfusion processes enabling highly specific and fast model calibration was developed. Harvest libraries consisting of frozen harvest samples from multiple CHO cell culture bioreactors collected at different process times were established. Model calibration was subsequently performed in an offline setup using a flow cell by spiking process harvest with glucose, raffinose, galactose, mannose, and fructose.

Major Results: In a screening phase, Raman spectroscopy was proven capable not only to distinguish sugars with similar chemical structures in perfusion harvest but also to quantify them independently in process-relevant concentrations. In a second phase, a robust and highly specific calibration model for simultaneous glucose (root mean square error prediction [RMSEP] = 0.32 g L^{-1}) and raffinose (RMSEP = 0.17 g L^{-1}) real-time monitoring was generated and verified in a third phase during a perfusion process.

Implication: The proposed novel offline calibration workflow allowed proper Raman peak decoupling, reduced calibration time from months down to days, and can be applied to other analytes of interest including lactate, ammonia, amino acids, or product titer.

KEYWORDS
flow cell, harvest library, model calibration, MVDA, Raman spectroscopy, spiking

Abbreviations: AOI, analyte of interest; DoE, design of experiments; GA, glucose addition; G_CS, glucose model for calibration set; mAb, monoclonal antibody; PC, principal component; PCA, principal component analysis; PLS, partial least squares; R^2, coefficients of determination calibration; RA, raffinose addition; R_CS, raffinose model for calibration set; R^2CV, coefficients of determination cross validation; RMSE, root mean square error; RMSECV, root mean square error cross validation; RMSEP, root mean square error prediction; R^2P, coefficients of determination prediction; RV, reactor volumes; SD, standard deviation; SNV, standard normal variate; VCD, viable cell density; VCV, viable cell volume.

1 | INTRODUCTION

Encouraged by regulatory authorities, a variety of different process analytical technology (PAT) tools have found their place in (bio)pharmaceutical manufacturing to ensure product quality in complex manufacturing processes.[1,2] By allowing continuous monitoring of process indicators (e.g., the glucose level in cell culture) connected to their corresponding control systems (e.g., glucose feed), PAT tools would improve the assurance of delivering a product with more consistent critical quality attributes. Spectroscopic techniques are characterized by their speed, non-invasiveness, and relatively low impact on the material under investigation.[3] Raman spectroscopy is arguably the most versatile technology for biotechnological upstream process (USP) monitoring given its low water interferences.[4,5]

The possibility of simultaneous monitoring of multiple process output variables and controlling them using multivariate models in conjunction with feedback loops has been demonstrated by several groups. Calibration models were generated to monitor metabolites (glucose, lactate, ammonia), amino acids, as well as product titer, viable cell density (VCD), viability, and even product quality attributes such as glycosylation.[6-11] Most of these models are exclusively based on *in-situ* collected Raman spectra aligned with reference analytics of process samples trying to capture process variability. However, many media components change concomitantly during the process, leading to difficulties generating chemometric models that are truly based on the fingerprint of the analyte of interest (AOI). This can lead to unawareness about the specificity of the generated models and predictions based on process evolution rather than compound-specific Raman bands. The importance to decouple Raman signals to obtain compound specific calibration models was discussed previously, where specific wavenumber regions assigned to glucose and lactate increased model reliability.[12] Furthermore, the model generation by alignment of offline measurements to simultaneously measured Raman spectra is limited to the offline analytics available on the respective site. Thus, Raman model generation for non-routinely measured compounds which might be interesting for developmental activities is difficult to realize and would require changes in the analytical pipeline.

Integrating sufficient analyte variability into the Raman calibration workflow remains one of the biggest challenges to generate robust calibration models. In general, analyte variability can either be integrated into the calibration process by executing a design of experiments (DoE) strategy incorporating variability to a set of bioreactor runs, or by spiking analytes of interest directly into the bioreactor to change their concentrations independently of the process matrix. The DoE strategy with multiple bioreactor runs was successfully applied for monitoring glucose and lactate concentrations in a cross-scale experiment composed of 5, 200, and 2000 L bioreactor runs with variability introduced to 5 L bioreactor runs.[13] Furthermore, a miniatured bioreactor system allowing up to 48 simultaneous cultivations at different conditions was shown to facilitate Raman model calibration for reliable glucose and lactate predictions.[14] Spiking strategies to augment analyte variability in the calibration data set were demonstrated in situ,[12,15] at-line,[16] and offline.[17] In contrast to the standard calibration approach aligning process spectra to reference analytics, these two strategies allow controlled coverage of the design space to achieve sufficiently high analyte variability for robust Raman calibration which might be especially important for steady-state perfusion processes with inherently low variability. Despite the ability to generate important analyte variability for model generation, these two approaches are very resource intensive. Dedicated bioreactor runs operated at varying setpoints or reactors used for spiking studies need to be executed for Raman model generation, interfering with other ongoing process development activities.

Although most of the reported Raman models were acquired in situ by inserting the Raman probe into the bioreactor, in-line monitoring of perfusion harvest using a flow cell offers an interesting alternative, provided that the harvest is collected through a cell separation device (e.g., ATF or TFF systems). Albeit some cell-related information cannot be extracted when acquiring Raman signals from the filtered harvest, the absence of cells and gas bubbles in the flow cell might be advantageous in terms of signal sensitivity, robustness, and scalability. Models developed in a flow cell can be used independently of the process scale when installed in a bypass loop, facilitating Raman model transfer across different production scales. As such, in-line monitoring of product titer in perfusion harvest using a stainless-steel flow cell was used to optimize an integrated continuous chromatography step,[18] whereas others developed a more sophisticated flow cell intensifying Raman signals to detect product breakthrough in a continuous protein A capture step.[19]

N-glycosylation is considered a major CQA of therapeutic monoclonal antibodies (mAbs) due to its potential effect on the efficiency and safety profile of the product. The glycosylation pattern of recombinant proteins is highly dependent on the cultivation conditions that need to be well controlled to deliver a consistent product quality.[20] Hence, factors that impact glycosylation need to be monitored and controlled ideally throughout the manufacturing process.[21] To modulate glycosylation patterns during a production process or to match glycosylation patterns to a reference product, feeding sugars such as mannose, galactose, fructose, or raffinose has been proven an effective way to reach target specifications.[22–27] Direct control with known modulators is of particular interest in quasi-steady-state continuous perfusion processes, where product quality attributes are kept constant throughout the process.

The goal of this study was to develop a systematic Raman calibration workflow to generate highly specific Raman models for perfusion cell culture with inherently low process variability and thereby overcoming limitations of current calibration approaches such as a lack of analyte variability in the calibration data, unsatisfactory analyte specificity due to correlation with simultaneously changing process compounds, or resource and time intensive calibration activities. The novel calibration workflow consisting of a harvest library containing process samples from multiple bioreactor runs covering different process time points enabled rapid model generation by spiking without impacting ongoing bioreactor runs. Model calibration was performed in an offline setup using a flow cell which was finally connected to the perfusion process harvest for real-time predictions. The obtained

results show that a systematic spiking approach can build a robust basis for the training of chemometric models particularly for substances having similar chemical structures in complex matrices. This approach not only speeds up a model generation for perfusion cultures from several months down to weeks or even days but more importantly allows breaking correlations of the batch evolution with the analytes of interest.

2 | EXPERIMENTAL SECTION

2.1 | Perfusion culture process, monitoring, and control

A proprietary CHO-K1 cell line producing a bispecific mAb was expanded in an incubator (Multitron 4, Infors HT, Bottmingen, Switzerland) for 21 days using a proprietary chemically defined medium (Merck Serono SA, Corsier-sur-Vevey, Switzerland). Perfusion bioreactors (Labfors 5 Cell, Infors HT, Bottmingen, Switzerland) were inoculated with a seeding density of 0.6×10^6 viable cells mL^{-1}. Culture conditions were maintained at 36.5°C with a DO setpoint of 50% (VisiFerm DO Arc, Hamilton, Bonaduz, Switzerland). The pH was controlled at 7.07 ± 0.17 (EasyFerm Plus Arc, Hamilton, Bonaduz, Switzerland) by CO_2 sparging and a 1.1M Na_2CO_3 solution. Bioreactors were operated at 2 L working volume and perfusion was started on day 0 and kept constant at 1.3 reactor volumes per day (RV/day). Bioreactor harvests were gravimetrically controlled to maintain the bioreactor weight constant using either alternating tangential flow filtration (ATF2H, Repligen, Waltham, MA) or tangential flow filtration (Levitronix technologies Inc, Framingham, MA) with polyether sulfone hollow fibers having a pore size of 0.22 μm (Repligen). After an initial growth phase, an online capacitance probe (Incyte Arc, Hamilton, Bonaduz, Switzerland) was used to keep the viable cell volume (VCV) constant at 12%.

2.2 | Hardware setup and data acquisition

A custom-made flow cell from stainless steel 316 L with a chamber volume of 0.95 mL was used for all Raman measurements. The flow cell allowed to insert any commercially available bioreactor probe compatible with a bioreactor headplate PG 13.5 thread. The outlet of the chamber was placed at the highest point to prevent bubble formation that could impair spectral acquisition. More details about the flow cell can be found in the Supporting Information section (Supporting Information Figures S1 and S2). A Raman spectrometer Multispec Raman (tec5 AG, Steinbach, Germany) using the software MultiSpec Pro II (tec5 AG, Steinbach, Germany) with 785 nm laser excitation wavelength and 500 mW power output was used in combination with the InPhotonics RamanProbe (InPhotonics, Norwood, MA) for the initial screening phase using spectral acquisition time of 6 × 20 s. For the model calibration and verification, 12 × 20 s scans were performed using a Raman Immersible Probe MSR M571 (tec5 AG, Steinbach, Germany) with a ViewPort PG13.5 sensor interface (SCHOTT AG, Landshut, Germany). In the offline setup, light-protected and tempered harvest (37°C) was pumped through the flow cell at a flow of 2 L day^{-1} matching the approximate harvest speed when connected to the perfusion bioreactor, resulting in an average residence time of 40 s.

2.3 | Chemometric modeling procedure

Data analysis was performed using The Unscrambler X (CAMO Software, Oslo, Norway). A standard spectral preprocessing procedure was applied for all chemometric modeling activities. Spectra were pre-treated by a 1st derivative with Savitzky-Golay filter (2nd order polynomial, 31 points window) followed by standard normal variate (SNV). Preprocessed spectra were then truncated to wavenumbers from 450 to 1800 cm^{-1} and used for further principal component analysis (PCA) and partial least squares (PLS) regression. The number of latent variables was determined by using the standard approach of The Unscrambler X software unless otherwise mentioned. Performance of the models was compared based on the root mean square error (RMSE) of the calibration dataset, five-fold cross-validation (root mean square error cross validation [RMSECV]) and prediction (root mean square error prediction [RMSEP]), and the corresponding coefficients of determination coefficients of determination calibration (R^2), coefficients of determination cross validation (R^2CV), and coefficients of determination prediction (R^2P) as previously described.[12,19]

2.4 | Reference analytics

Glucose concentrations and other CPPs such as lactate, ammonia, glutamine, glutamate, VCD, and viability were measured using the Bioprofile Flex2 (Nova Biomedical, Waltham, MA). The supplier indicated an analytical standard deviation (SD) for glucose of 0.07 g L^{-1}. VCV was calculated as follows:[28]

$$\mathrm{VCV} = \frac{\frac{4}{3} \times \pi \times \left(\frac{D}{2}\right)^3 \times \mathrm{VCD}}{V} \times 100 \quad (1)$$

where D is the average cell diameter, VCD the viable cell density, and V is the cell culture volume, assuming a spherical shape of the cells. Raffinose was quantified on an Agilent 1200 HPLC system combined with a refractive index detector (Agilent RID, G1362A). Compounds were separated using a Hiplex H column (250 × 4.6 mm, PL1170-6830, Agilent Technologies AG, Basel, Switzerland) with a guard column Hiplex H column (3 × 5 mm, PL1670-0830, Agilent Technologies AG, Basel, Switzerland) as described in the literature.[29] The flow rate was adapted to 0.2 mL min^{-1} and the analytical SD was determined to be 0.04 g L^{-1}. Due to the absence of raffinose in the platform perfusion medium, raffinose concentrations during the verification run on day 1 and from day 8–13 were assumed to be 0 g L^{-1}.

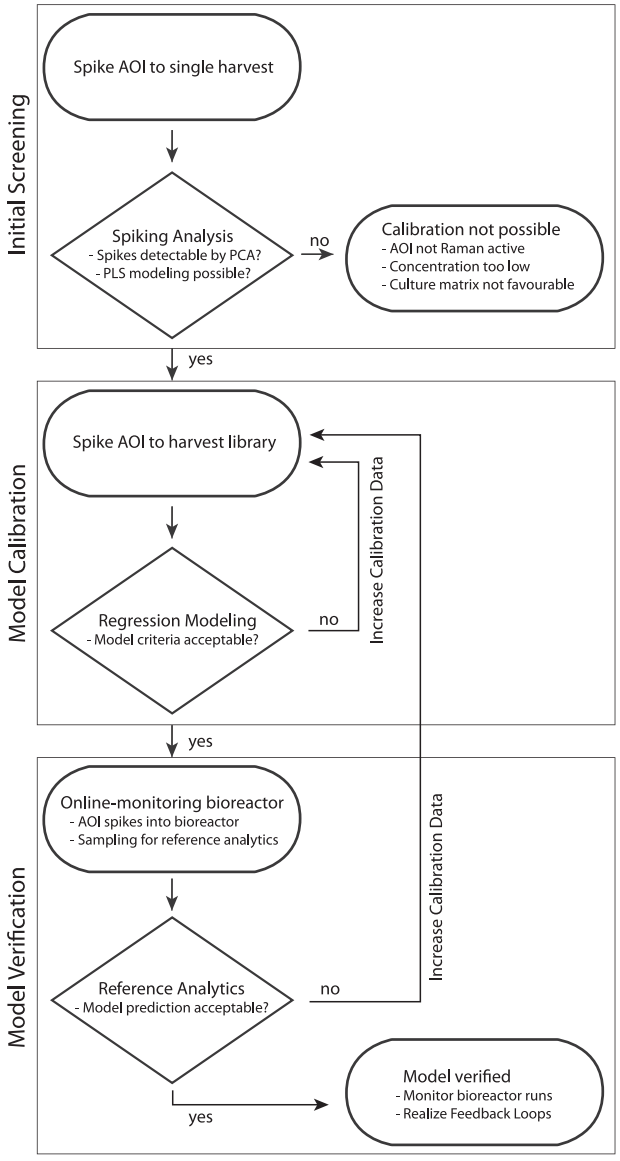

FIGURE 1 Spiking chart for the spiking-based Raman calibration workflow. The workflow consists of three phases. Initial screening phase: identification if calibration of a certain analyte is possible. Model calibration phase: actual model calibration using the harvest library. Model verification phase: final stage to verify calibration models with independent bioreactor runs

2.5 | Experimental workflow for Raman model calibration

An experimental workflow consisting of three main steps was developed to ensure robust Raman sensor calibration (Figure 1):

Initial screening phase: Initial screening to assess whether analytes of interest can be distinguished by Raman in the desired concentration range and process matrix.

Model calibration phase: Data acquisition of the AOI for the model calibration using a harvest library.

Model verification phase: Model verification in a perfusion bioreactor run.

During the initial screening phase, a full factorial design with two factors (glucose and one out of the four secondary sugars) with five levels was performed and prepared in duplicates, leading to 50 spectra per glucose-secondary sugar combination. As there were four different secondary sugars, a total of 200 spectra resulted in a combined calibration data set. Furthermore, one extra solution of each corner condition and three extra central point solutions were prepared as a test set, resulting in 28 spectra. Glucose levels were equally distributed within a range of 3.45–13.45 g L^{-1}. The concentration ranges of secondary sugars were chosen according to the values found to modulate glycosylation patterns in the literature.[22,26] This resulted in 0–4 g L^{-1} for galactose, mannose, and fructose, and in 0–10 g L^{-1} for raffinose all equally distributed on five levels. Glucose (glucose-monohydrate for microbiology, CarlRoth, Karlsruhe, Germany), raffinose (D(+)raffinose pentahydrate, Apollo scientific, Bredbury, UK), galactose (D(+)galactose 99+%, Acros organics, Thermo Fisher Scientific, Waltham, MA), mannose (D(+)mannose 99+%, Acros organics, Thermo Fisher Scientific), and fructose (D(−)fructose, Fisher Scientific, Thermo Fisher Scientific) were weighted and dissolved in 15 mL perfusion harvest for spectral acquisition in the spiking experiments. The same perfusion harvest was used for all the spikes in the initial screening phase to keep the harvest matrix constant.

For the model calibration phase, harvest samples of five different perfusion bioreactor runs, each sampled on five different days to capture intra- as well as inter-process variability were used (Table 2). Harvests of bioreactor runs R1 and R2 were sampled and directly spiked for model calibration in the flow cell setup. Harvest samples of the three remaining bioreactors (R3, R4, and R5) were light-protected and stored at −80°C thereby building a harvest library. Harvest library samples were thawed and handled in the same way as the fresh harvest once the model calibration was performed (Figure 2). The variability of the main process parameters was assessed in all the harvests using a Bioprofile Flex2 (Nova Biomedical). Spiking for raffinose ranged from 0 to 10 g L^{-1}, for glucose from 3.3 to 12.7 g L^{-1}, and is summarized in Figure 3B.

Model verification was performed by connecting the flow cell to the harvest stream of a perfusion bioreactor run and continuous spectra acquisition was performed for 13 days during the stable operation phase. Two raffinose additions (RAs, RA1 and RA2) followed by two glucose additions (GAs, GA1 and GA2) were performed directly into the bioreactor to assess the model prediction performance of the model over the entire calibration range. Raffinose was fed for 4 h with a flow of 2.6 L day^{-1} by exchanging the standard perfusion medium with a medium supplemented with 50 g L^{-1} raffinose. Considering an average VCV of 12.2% during the model verification phase (data not shown) and washing out of the reactor already during the addition phase, a theoretical raffinose peak concentration of 10.9 g L^{-1} for RA1 was calculated based on the bioreactor mass balance (see Supporting Information). For RA2 with a starting concentration of 0.2 g L^{-1} raffinose determined by reference HPLC analytics, a peak concentration of 11.1 g L^{-1} was calculated. Glucose alterations were initiated by adding bolus shots of a 300 g L^{-1} glucose feed. A bolus addition of 50 and 12 mL was executed for GA1 and GA2, respectively. Glucose peak concentrations were

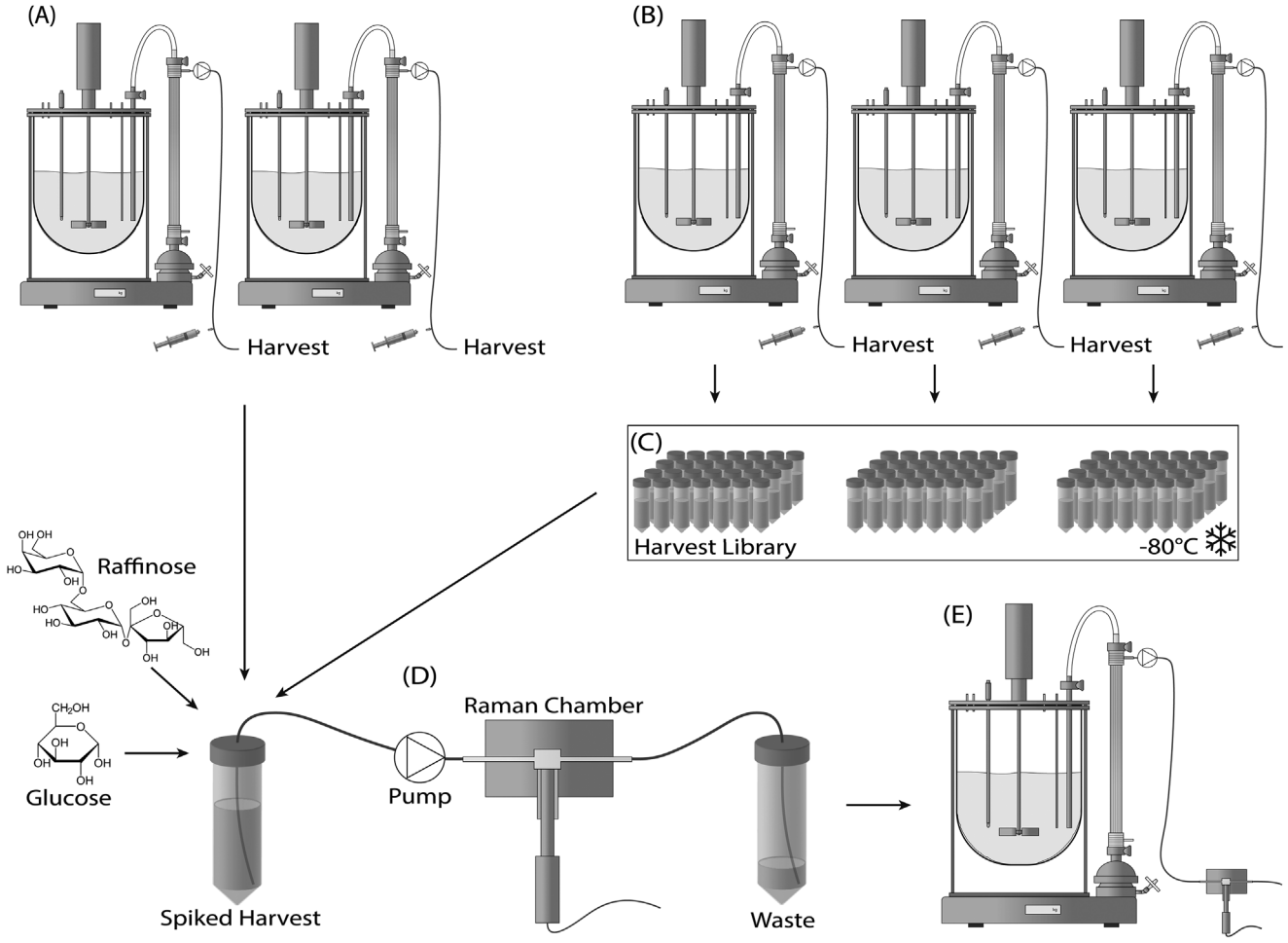

FIGURE 2 Schematic of the harvest library spiking approach. Fresh harvest samples of two reactors (A) and harvest samples of three reactors (B) which were frozen and stored in a harvest library (C) were used for model calibration. Harvests were spiked with glucose and/or raffinose and pumped through a Raman flow cell (D) to generate Raman calibration data. The flow cell was then connected directly to the bioreactor run (E) for model verification

calculated to be 11.8 g L^{-1} (GA1) and 5.6 g L^{-1} (GA2) considering starting concentrations of 3.2 and 3.6 g L^{-1}, respectively. The bioreactor harvest stream was sampled during alterations for raffinose and glucose reference analytics.

3 | RESULTS AND DISCUSSION

3.1 | Initial screening phase

The purpose of the initial screening phase is to determine whether the different analytes of interest can be distinguished by Raman spectroscopy in a best-case scenario without process disturbances, fluorescence, or other detrimental effects. By using a single harvest for all initial spiking experiments, sugar concentrations could independently be varied without changing the matrix at the same time.

PCA was performed after the standard spectral pretreatment showing distinct clustering of the Raman spectra with spiked sugars. The glucose variability in the spectra is mainly captured by principal component (PC)1 and PC2 (Figure 4B), whereas variability introduced by other sugars can be explained by PC3-5(Figure 4A). All spectra containing only a certain glucose concentration but without additional sugar were populating in the center of the PCA plot. With increasing concentrations of the secondary sugar, the recorded spectra shifted into a specific direction away from the common center, the further away, the higher the added concentration of the secondary sugar. This showed not only that the investigated sugars can be differentiated but can also be quantified within the respective concentration range.

PLS modeling results underlined the feasibility of distinguishing and quantifying the investigated sugars within the perfusion harvest matrix (Table 1). In general, the models featured low RMSE, RMSECV, and RMSEP with R^2, R^2CV, and R^2P $\geq$ 0.95. To identify important wavenumber regions for the different PLS models, min–max normalized regression coefficients were used to color the background of the Raman spectra for the respective sugar models (Figure 5). Model regression coefficient patterns vary strongly between different sugars. As a first derivative was applied to the spectra during spectral pre-treatment, peak assignment to functional groups/bonds becomes

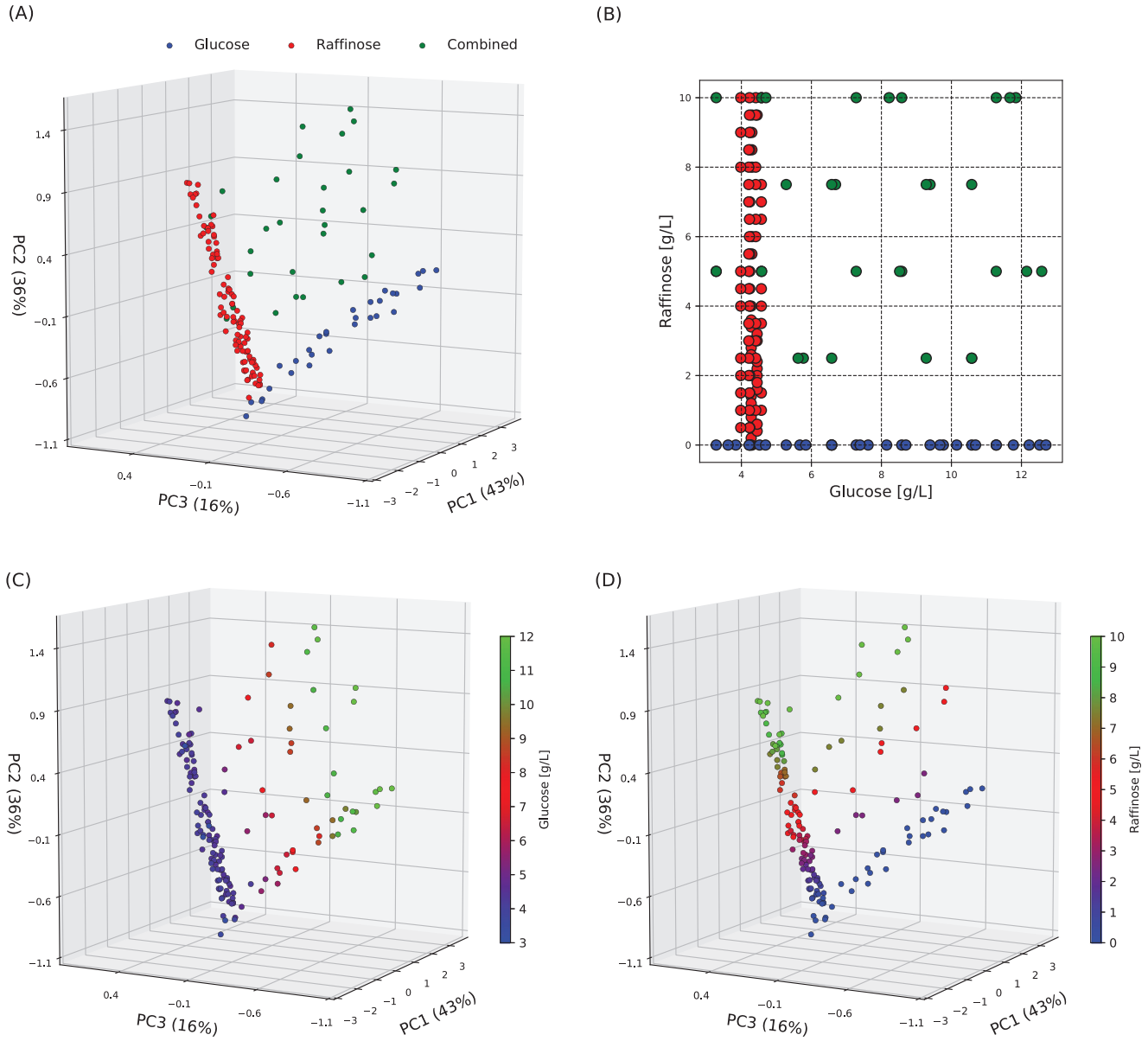

FIGURE 3 PCA plot colored by the spiked compounds are depicted in (A) and the corresponding theoretical concentrations in (B). Single raffinose spikes are colored in red, single glucose spikes in blue and mixed spikes in green. CS1 is composed of red samples, CS2 of red and blue samples, CS3 of red, blue, and green samples. PCA plots of the individual sugar concentrations are depicted for glucose in (C) and raffinose concentration in (D). CS, calibration set; PCA, principal component analysis

a challenging task and would go beyond the scope of this work. Nevertheless, the most important regions according to the regression coefficients are in good agreement with specific bands reported in the literature.[30,31] The highest regression coefficient region for glucose lies between 1100 and 1150 cm^{-1}, which is likely the particular COH bending (δ(COH)) at 1128 cm^{-1}. Raffinose shows the highest regression coefficients between 820 and 860 cm^{-1}, relating to the COC stretching (ν(COC)) at 836 cm^{-1}. Galactose was reported to show a strong distinctive CCO stretching band (δ(CCO)) at 526 cm^{-1}, which corresponds well with the high regression coefficients between 510 and 550 cm^{-1}. Mannose showed particularly high regression coefficients between 460 and 510 cm^{-1}, covering an important mannose peak at 488 cm^{-1} found by others.[31] Fructose showed one particularly important wavenumber region between 610 and 650 cm^{-1}, reported to be a very particular OCO-stretching band (δ(OCO)) at 631 cm^{-1}. A clear color pattern of the spectra according to the concentration is visible in the regions with important regression coefficients for all measured sugars, which can be used as another indication for a compound-specific calibration based on the molecular fingerprint.[12]

Overall, the specific sugar clustering seen in the PCA plot, successful PLS modeling with low RMSEP, and the assignment of the most important model regions to molecular vibrations indicated that a Raman calibration for the various sugars in the harvest matrix is possible.

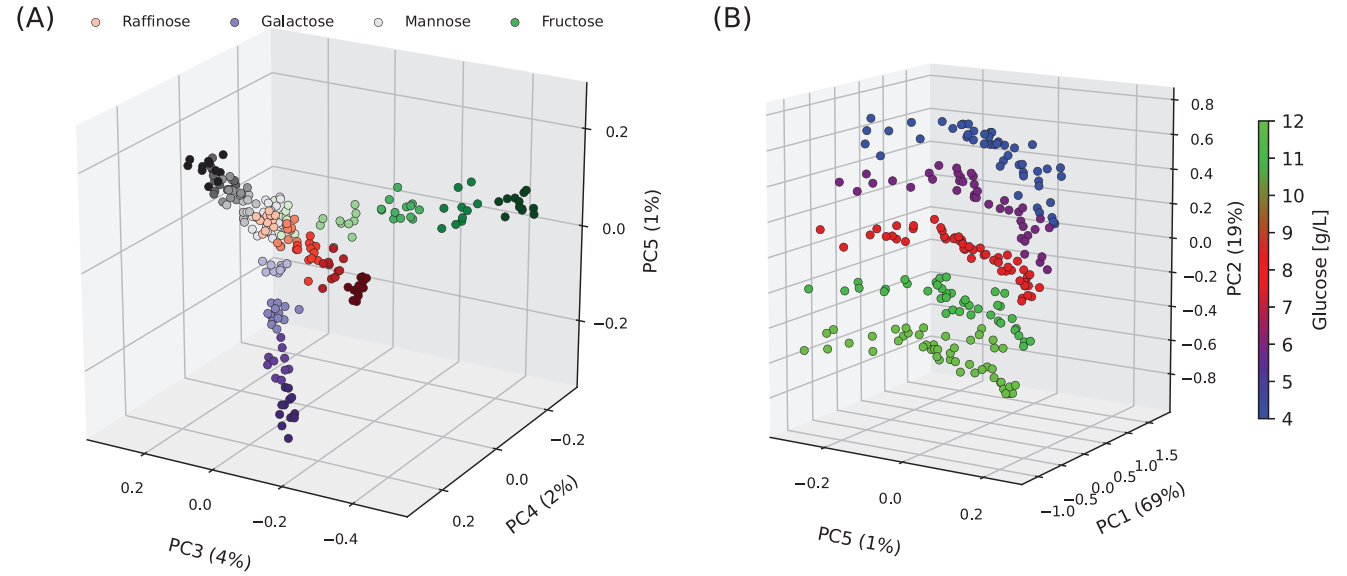

FIGURE 4 Principal component analysis (PCA) plots for the initial screening phase for raffinose, galactose, mannose, and fructose with increasing color intensity according to their concentration (A) and glucose (B). Originating from a common center point, all sugars populate in a distinct direction. The higher the concentration, the further away from the center, demonstrating that the Raman spectra can distinguish the various sugars in a concentration-dependent manner

3.2 | Model calibration phase

Since glucose is one of the main energy sources in mammalian cell culture, the model for secondary sugars should be reliable also in the presence of varying glucose concentrations. Raffinose is a trisaccharide containing a glucose molecule and was therefore chosen as a worst-case scenario to demonstrate the applicability of the calibration workflow for similar compounds. Galactose, mannose, or fructose exhibited similar strong differences from glucose in the initial screening phase, and thus similar performances can be expected.

Model calibration was based on harvests collected from five different perfusion bioreactor runs once they reached a stable operation. Each reactor was sampled on five different working days to create a diverse set of harvest matrices. In contrast to the initial screening phase, where only a single harvest with a particular matrix composition was used, the model calibration data set is based on spiking experiments with inherent process variability (samples from different process working days) but also inter-process variability (different bioreactor runs). Nutrient and metabolite concentrations as well as cell-specific parameter ranges of the calibration run harvests showed variability as expected from 12 reference runs throughout the steady-state perfusion phase (Table 2), which is a pre-requisite for obtaining robust Raman models. The means and SDs of the calibration and reference runs are close, providing a solid basis for spiking studies for the five calibration runs.

Harvest samples were either directly spiked and used for model calibration or frozen for later use. The spiking scheme in Figure 3B summarizes which concentrations of glucose and raffinose were present after the spikes were added. Different spiking strategies were assessed to see whether raffinose spiking alone (red points) leads to a robust calibration set (CS1), or if the data set should be augmented (CS2) by glucose spiking (blue points). A third calibration set (CS3) was built

TABLE 1 Modeling results for the initial screening phase

Sugar	Concentration range [g L^{-1}]	Suggested LV	RMSE [g L^{-1}]	RMSECV [g L^{-1}]	RMSEP [g L^{-1}]	R^2	R^2CV	R^2P
Glucose	3.45–13.5	3	0.45	0.46	0.41	0.984	0.983	0.988
Raffinose	0–10	5	0.28	0.31	0.29	0.990	0.988	0.990
Galactose	0–4	5	0.20	0.21	0.21	0.969	0.965	0.967
Mannose	0–4	5	0.21	0.22	0.26	0.966	0.962	0.948
Fructose	0–4	5	0.14	0.14	0.16	0.985	0.983	0.982

PLS models were generated for the five different sugars within the specified concentration ranges. The calibration data set consisted of 200 samples, and the test data set of 28 samples. A five-fold cross-validation was performed. LV, latent variables; R^2, coefficients of determination calibration; R^2CV, coefficients of determination cross validation; RMSE, root mean square error; RMSECV, root mean square error cross validation; RMSEP, root mean square error prediction; R^2P, coefficients of determination prediction.

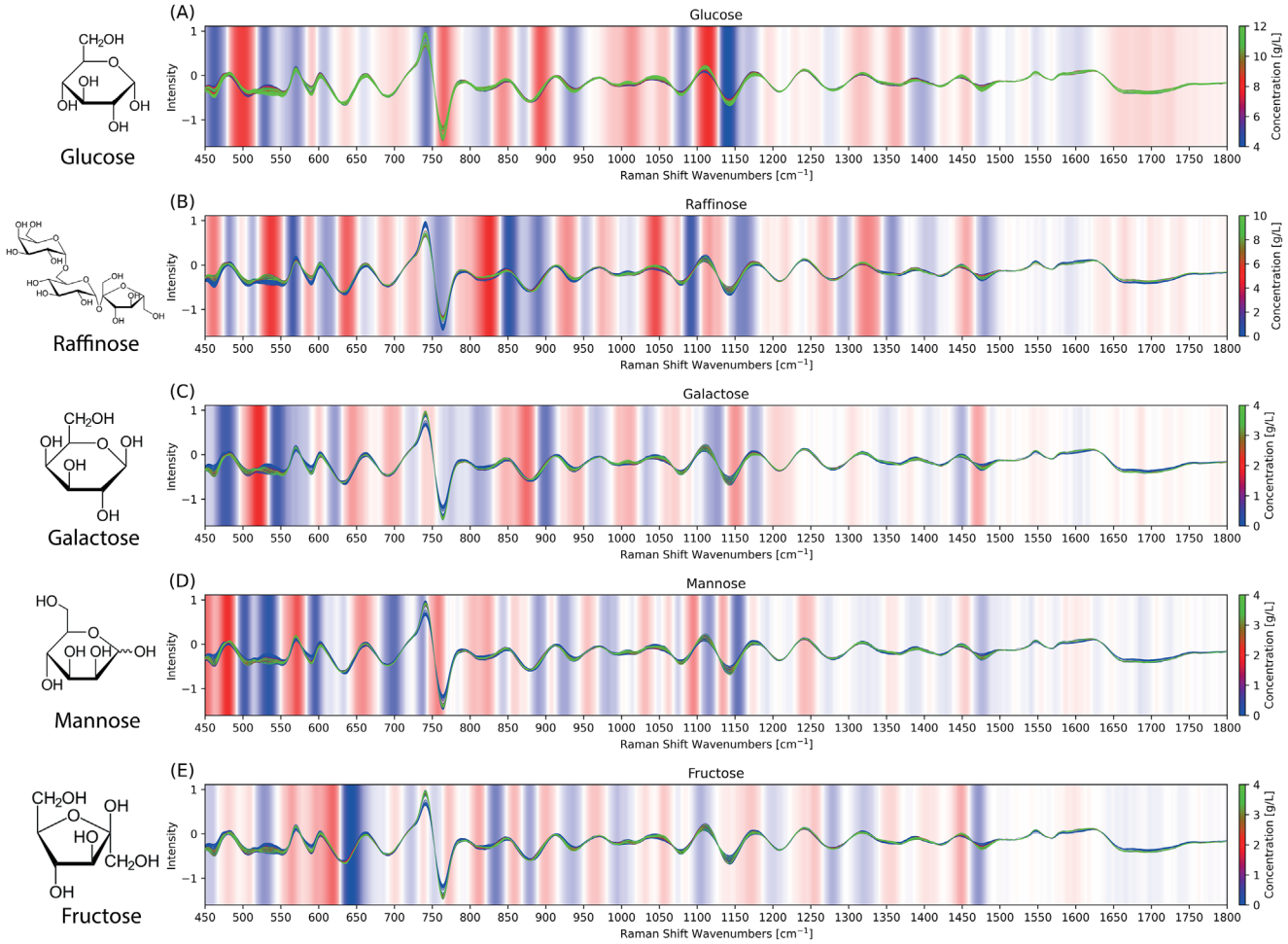

FIGURE 5 Raman spectra for glucose (A), raffinose (B), galactose (C), mannose (D), and fructose (E) in perfusion process harvest. All spectra were pretreated with a Savitzky–Golay filter (derivative 1, polynomial = 2, window size = 31) followed by standard normal variate (SNV). Background bands represent the min-max normalized regression coefficient for the respective wavenumber for the individual models. Blue colors show the negative contribution of the wavenumber to the predicted value, whereas red colors indicate a positive contribution. The more intense the coloring, the more important the wavenumber for the respective model

including both the raffinose spikes (red points) and glucose spikes (blue points), but also all the combinations of the two sugars (green points) present in the measured harvest at the same time. As different harvests contained varying starting concentrations of glucose, measurement points with equal amounts of spiked glucose and raffinose scatter slightly along the glucose axis and do not perfectly overlap as one would maybe expect of a DoE approach. The theoretical spiking scheme depicted in Figure 3B could almost identically be projected by the PCA plot Figure 3A, just with a small rotation, showing that glucose and raffinose influence the Raman spectra in an independent yet concentration-dependent manner. Raman spectra were then colored according to the glucose concentration (Figure 3C) and according to the raffinose concentration (Figure 3D), revealing a distinct separation along the concentration gradient. Already three PCs cover the variability introduced by the spiking approach. A PCA plot colored according to the reactor run origin of the spiked harvest can be found in the Supporting Information (Figure S3).

PLS regression was performed for the three calibration sets, both for glucose and raffinose (Table 3). All three models for glucose feature low RMSE ≤ 0.36 g L^{-1} and RMSECV ≤ 0.38 g L^{-1}. Model glucose model for calibration set (G_CS)1 showed even smaller RMSE and RMSECV of 0.08 and 0.09 g L^{-1}, however, showed rather low R^2 and R^2CV with 0.741 and 0.698. For the other two glucose models, R^2 and R^2CV were above 0.975. A potential explanation for the low R^2 values for G_CS1 is the limited glucose range used for the model calibration of only 3.3–4.7 g L^{-1}. For this reason, the model might not be able to distinguish the free glucose from the glucose molecule present in the raffinose molecule as peaks are not ideally decoupled. Despite the low RMSE and RMSECV of model G_CS1, the PLS results must be taken with care and proper verification data should be considered. Raffinose models showed low RMSE $\leq$ 0.29 g L^{-1} and RMSECV $\leq$ 0.30 g L^{-1} with all R^2 and R^2CV $\geq$ 0.992. Again, the model containing only raffinose spikes (raffinose model for calibration set [R_CS]1) showed the lowest RMSE and RMSECV, however with high R^2 values. In general, the

TABLE 2 Perfusion harvest composition for model calibration

Data origin	Statistics	Glutamine [mM]	Glutamate [mM]	Glucose [g L^{-1}]	Lactate [g L^{-1}]	Ammonia [mM]	VCD [10^6 cells mL^{-1}]	Viability [%]
Cal. Runs	Range	0.66–1.3	1.23–2.2	3.3–4.7	0.03–0.61	2.31–5.94	19.2–29.8	68–96
Ref. Runs	Range	0.41–1.39	0.37–1.96	3.1–5.8	0–0.48	1.85–5.75	14.9–31.8	92–99
Cal. Runs	Mean	0.99	1.69	4.10	0.16	4.19	24.1	89
Ref. Runs	Mean	0.88	1.34	4.71	0.16	3.17	20.5	98
Cal. Runs	SD	0.19	0.22	0.41	0.22	1.18	2.6	8
Ref. Runs	SD	0.13	0.25	0.47	0.11	1.00	1.5	1

Overview of nutrient compositions, metabolite concentrations, and cell parameters of the five runs (harvest library) used for Raman model calibration and of 12 reference runs. Ranges from the harvest library, mean and SD represent inter-process statistics for the respective data origin. SD, standard deviation; VCD, viable cell density.

TABLE 3 Modeling results for final glucose and raffinose models

Name	Sugar	Cal. set	Pred. set	LV	RMSE [g L^{-1}]	RMSECV [g L^{-1}]	RMSEP [g L^{-1}]	R^2	R^2CV	R^2P
G_CS1	Gluc	95	18	3	0.08	0.09	1.59	0.741	0.698	0.565
G_CS2	Gluc	125	18	2	0.31	0.33	0.39	0.978	0.977	0.973
G_CS3	Gluc	155	18	2	0.36	0.38	0.32	0.979	0.977	0.982
R_CS1	Raff	95	23	3	0.18	0.18	0.84	0.997	0.996	0.948
R_CS2	Raff	125	23	3[a]	0.22	0.22	0.17	0.996	0.995	0.998
R_CS3	Raff	155	23	3	0.29	0.30	0.25	0.993	0.992	0.995

Models based on calibration sets CS1, CS2, and CS3 for glucose (G) and raffinose (R) were evaluated on their prediction performance in the verification run. Spectra with aligned reference analytics for glucose and raffinose served as a prediction set. RMSEP versus LVs plots for all models can be found in the Supporting Information section (Supporting Information Figure S4). LV, latent variables; R^2, coefficients of determination calibration; R^2CV, coefficients of determination cross validation; RMSE, root mean square error; RMSECV, root mean square error cross validation; RMSEP, root mean square error prediction; R^2P, coefficients of determination prediction.

[a] 3 LVs were used for the sake of comparison. RMSEP as a function of LVs are provided in the Supporting Information (Figure S4).

obtained RMSE for glucose are in a similar range as models reported in the literature,[6,9–11] even in the presence of a secondary sugar like raffinose.

To assess whether concentration changes of untracked harvest component or freeze-thawing of harvest library samples (Run 3, 4, and 5) showed different concentration predictability compared to freshly spiked harvest samples (Run 1 and 2), a leave-one-out regression analysis was performed based on the largest calibration set CS3 (Supporting Information Table S1). Therefore, one of the five runs was removed from the calibration set and used as a test set. This was iteratively done for all five runs leading to predictions for glucose and raffinose. The resulting RMSEP for glucose and raffinose does not drastically deviate from the overall model based on all five runs for CS3, therefore underlining that the five chosen runs indeed cover process variability well and the harvest library approach is legitimate for Raman model generation.

3.3 | Model verification phase

The initial screening phase and the model calibration phase were performed in an offline setup, meaning the flow cell was not connected directly to a bioreactor harvest stream. For the model verification phase, the flow cell was connected to the harvest stream of a perfusion bioreactor run and online predictions were made for glucose and raffinose (Figure 6) using the six models built in the model calibration phase (Table 3). Due to the absence of raffinose in the standard perfusion medium, two RAs into the bioreactor of 10.9 g L^{-1} were made on day 1(RA1) and 11.1 g L^{-1} on day 4 (RA2) to assess the model performance on a wide concentration range. Furthermore, two GAs were performed on day 10 (GA1) and day 12 (GA2) to investigate the raffinose model performance in the presence of changing glucose concentrations. The GA1 addition targeted a bioreactor concentration of 11.8 g L^{-1} glucose to simulate a very strong glucose deviation. The addition of GA2 on the other hand simulated smaller glucose drift up to 5.6 g L^{-1} being more realistic for steady-state perfusion process fluctuations.

Raffinose predictions based on CS1 (Figure 6A) led to a reasonable performance on RA1 and RA2, even though the target peak concentrations of raffinose were slightly underpredicted. After stopping RA, a slow decrease of the raffinose concentration due to consumption and washing out by the perfusion process was measured. Glucose predictions during the raffinose alterations are in close alignment with the reference points, showing no interference of the raffinose with the glucose prediction. However, strong prediction deviations for the glucose and raffinose models (G_CS1 and R_CS1) occurred when glucose was added at 11.8 g L^{-1} (GA1). These models are not able to distinguish raffinose from glucose, leading to a prediction of up to 3 g L^{-1} raffinose when there was no raffinose present at all. Glucose on the other hand was strongly underpredicted with 6.4 g L^{-1} instead of 11.8 g L^{-1}. Even for the smaller GA2 of 5.6 g L^{-1}, a similar behavior was observed, resulting in an RMSEP for glucose of 1.59 g L^{-1} and raffinose of 0.84 g L^{-1} (Table 3). Glucose variability during the process was not sufficient to properly train the raffinose model, clearly demonstrating a need to extend the calibration range to make the model more robust in case of glucose fluctuations during the process. Proper model verification by performing compound spiking into the verification run should therefore routinely be done to assess model accuracy, robustness, and reliability as PLS regression metrics alone proved to be insufficient.

Raffinose models based on CS2 (Figure 6B) showed good predictive power for the RAs RA1 and RA2 throughout the whole run. Only a marginal drop of 0.1–0.2 g L^{-1} raffinose prediction was detected during the large GA GA1. No impact on raffinose prediction for GA2 was observed, simulating a realistic glucose deviation during the process. Thus, the raffinose model R_CS2 can be considered a specific and robust model for process monitoring. The glucose prediction by G_CS2 during the raffinose spikes showed small increases in the glucose prediction of about 0.2–0.3 g L^{-1}, indicating some leftover correlation between the two sugars. Nevertheless, glucose predictions by G_CS2 were in good alignment with the target concentration and the reference analytics during GAs GA1 and GA2. An overall low RMSEP of 0.39 g L^{-1} for glucose and 0.17 g L^{-1} for raffinose resulted in CS2 (Table 3).

The prediction model for raffinose R_CS3 based on the largest calibration set CS3 including combination spikes in the calibration data showed good alignment with reference analytics (Figure 6C). Compared to R_CS2, this model led to a slightly higher RMSEP of 0.25 g L^{-1}, meaning that combination spiking did not bring any significant benefit during the perfusion run verification. Combination spiking however had a positive impact on the glucose predictions with an RMSEP of 0.32 g L^{-1}, even artifacts of raffinose during RAs RA1 and RA2 could be removed.

Comparing regression coefficients of the glucose model G_CS1, which was generated by simply aligning offline glucose measurements with Raman spectra, to the regression coefficients of the glucose model G_CS2 and G_CS3, clear differences in the absolute values are visible (Figure 7A). The regression coefficient of the G_CS1 model show not only much lower absolute values, but peaks are less defined and noisier, being a sign of weak model calibration. This demonstrated that the G_CS1 model had not enough glucose variability in the calibration data to generate a robust and specific model for glucose. Albeit the overall regression coefficients are similar in structure for all raffinose models, the numeric contributions at the fingerprint regions are slightly more pronounced for R_CS2 and R_CS3 (Figure 7B). The largest difference distinguishing R_CS1 from the other models (R_CS2 and R_CS3) arises between 1100 and 1150 cm^{-1}. The more robust raffinose models R_CS2 and R_CS3 show much lower regression coefficients around 1128 cm^{-1}, being exactly the very particular glucose COH bending (δ(COH)) at 1128 cm^{-1}, and higher regression coefficients toward 1150 cm^{-1}. The decoupling of the Raman bands for glucose and raffinose in this particular region by glucose spiking is most likely responsible for the higher specificity of the Raman prediction and demonstrates the necessity of proper spiking studies to decouple Raman bands.

Overall, successful model verification demonstrated that the developed model calibration approach spiking fresh harvest or a harvest library showed significant advantages compared to the traditional

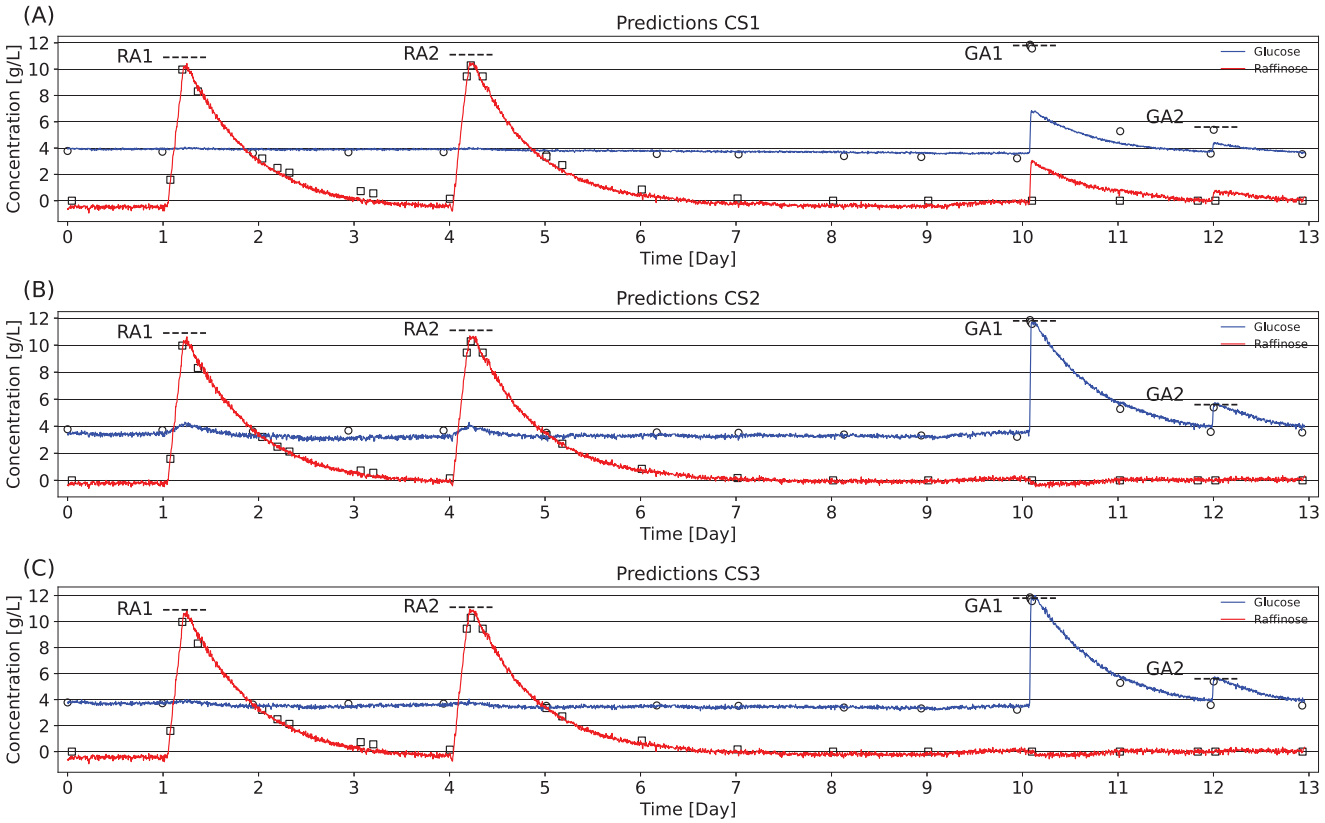

FIGURE 6 Raman predictions of the verification run using models based on calibration set CS1 (A), CS2 (B), and CS3 (C). RA1 and RA2 represent raffinose additions of 10.9 and 11.1 g L^{-1} to the bioreactor, whereas GA1 is a glucose addition of 11.8 g L^{-1} and GA2 a glucose addition of 5.6 g L^{-1}. Raw predictions without moving average are shown and aligned with reference values for glucose (circles) and raffinose (squares). CS, calibration set; GA, glucose addition; RA, raffinose addition

in-situ model calibration. Once a harvest library is established, new calibration models can be created at any time by thawing harvest and spiking it with analytes of interest. In this way, the Raman fingerprint of an analyte in a complex matrix is used to calibrate models while lowering matrix and process effects, constituting an alternative for perfusion cell culture to the relative resource and time-intensive DoE calibration strategies with parallelized bioreactor runs operated at varying process setpoints to induce sufficient process variability.[13,14] Although all three calibration sets CS1, CS2, and CS3 showed good raffinose predictions during the first 10 days without larger glucose fluctuation, significant prediction deviations occurred when challenged by glucose alterations for the models with no glucose spikes (G_CS1 and R_CS1). The inherent glucose variability of the different harvest samples was not sufficient for the PLS regression to distinguish glucose from raffinose, and spiking proved to be necessary to achieve proper peak decoupling to generate compound-specific model predictions. In contrast to successful previous spiking studies where only one or a few spikes were added directly to a set of bioreactor runs,[12,15] the proposed spiking approach for perfusion cell culture increases spiking capabilities and flexibility as spikes are performed externally of the bioreactor in a small harvest volume. Whether combination spiking is necessary, or if single compound spiking is sufficient as demonstrated by the raffinose model R_CS2, remains an open question and might depend on the respective compounds being modeled. With the increasing number of components to be spiked at the same time, the DoE approach might show an advantage of significantly reducing the number of experiments. With raffinose, it was demonstrated that calibration models for analytes absent in the standard perfusion process can be created offline without the need of adding the analyte to a single bioreactor.

The novel calibration approach has been applied to a steady-state perfusion process for an mAb producing CHO cell line to extend calibration ranges and improve the decoupling of process compounds. Dynamic perfusion cultures with the presence of a cell-free harvest stream would equally profit from this calibration approach. Whether harvest-based calibration data acquisitions are sufficient to enable robust generic models such as for different media and cell lines remains an open question. The rapidity of generating adequate Raman spectra using the spiked harvest library approach would certainly facilitate the development of process specific as well as general Raman models. Generating representative library samples for batch and fed-batch cultivations without a filtration device is more challenging due to the presence of cells. Thus, applying a similar design-of-experiment spiking approach to a fresh culture sample represents a promising method to decouple concomitant changing culture compounds.

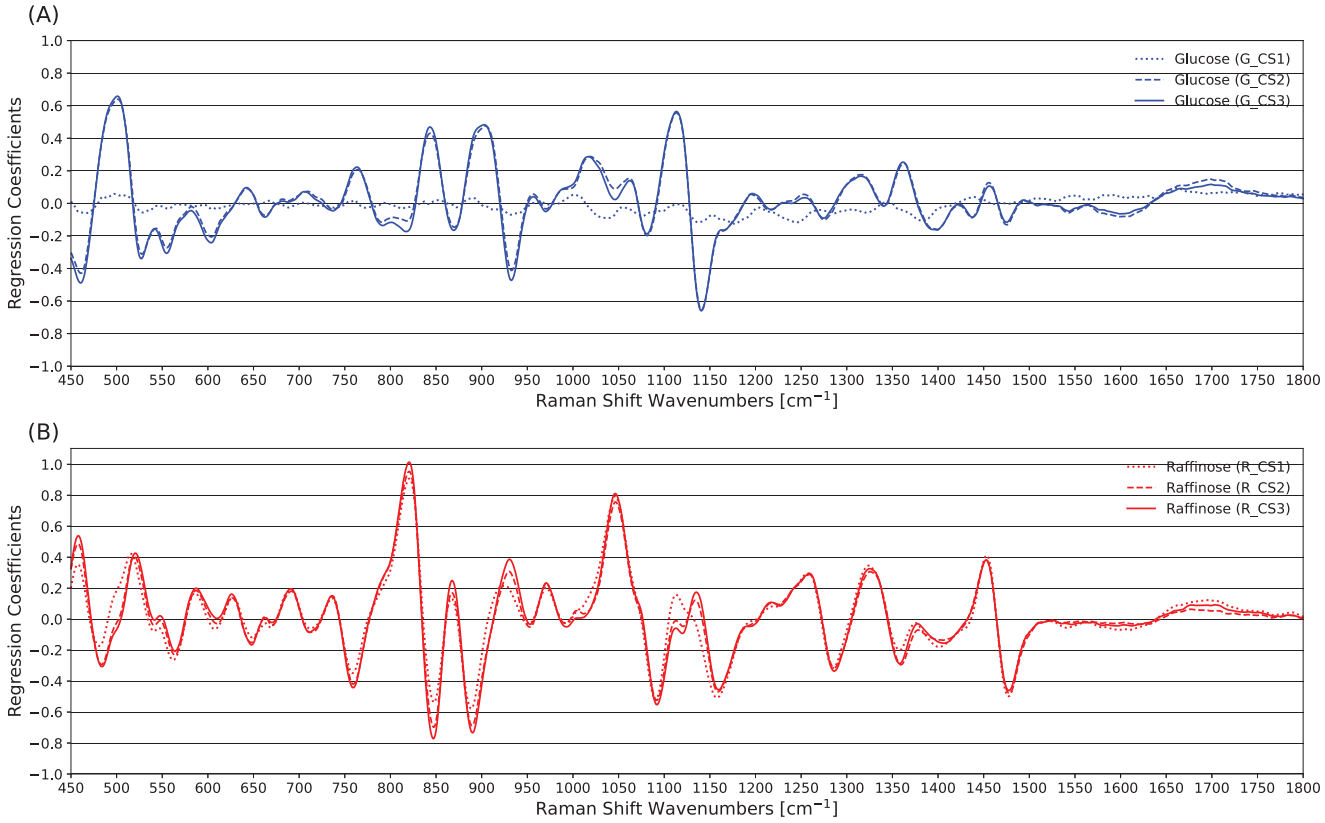

FIGURE 7 Regression coefficients for glucose models (A) and raffinose models (B)

4 | CONCLUSIONS AND OUTLOOK

In this work, a systematic spiking-based Raman calibration workflow using a harvest library was developed to calibrate chemometric models and to monitor analyte concentrations using a flow cell connected to the harvest stream of a perfusion process. In a showcase to generate models of molecularly similar compounds, well-decoupled calibration models for glucose and raffinose with an RMSEP of 0.32 and 0.17 g L^{-1}, respectively, were achieved.

Providing appropriate calibration data sets in place of observations around normal operating points is paramount for chemometric modeling, whether standard PLS regression or more advanced mathematical regression techniques are used. The developed calibration workflow can be applied to other nutrients or metabolites as long as their concentration is in a Raman detectable range, eliminating the need for complex and time-consuming offline data alignment.

Ultimately, the presented workflow may pave the way for advanced monitoring and control of a number of critical cell culture components that can have a significant impact on critical quality attributes including post-translation modifications such as glycosylation or glycation.

ACKNOWLEDGEMENTS

The authors would like to thank Pavel Dagorov and his workshop team (FHNW) for building the Raman flow cell, Martin Held and Tsvetan Kardashliev (D-BSSE, ETH Zürich) for support in HPLC sugar analysis, and Damian von Blarer (FHNW) for automation support. Christian Ott (SCHOTT AG), Sebastian Dedera, and Christian Lux (tec5 AG) kindly assisted with the Raman device setup. Moreover, the authors would like to acknowledge Merck Serono SA (an affiliate of Merck KGaA) for material support and specifically the Bioprocess Sciences (BPS) team for valuable discussions and support throughout the project, especially Martin Jordan, Golzar Mesbah, Loic Chappuis, Alexandre Chatelin, Arnaud Perilleux, Hervé Broly, and Jonathan Souquet. P.R.: Conceptualization, formal analysis, investigation, methodology, visualization, writing – original draft. J.K.: Data curation, formal analysis, investigation, methodology, visualization, writing – original draft. D.T.: Formal analysis, investigation. C.H.: Conceptualization, methodology, supervision, writing – review & editing. J.-M.R.B.: Conceptualization, investigation, project administration, resources, supervision, visualization, writing – review & editing. T.V.: Conceptualization-Supporting, Formal analysis-Supporting, Funding acquisition-Lead, Investigation-Supporting, Methodology-Supporting, Project administration-Lead, Resources-Lead, Supervision-Lead, Visualization-Equal, Writing – review & editing-Lead.

Open access funding provided by Fachhochschule Nordwestschweiz FHNW.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Patrick Romann https://orcid.org/0000-0002-7555-0598
Christoph Herwig https://orcid.org/0000-0003-2314-1458
Thomas K. Villiger https://orcid.org/0000-0003-0036-2522

LITERATURE

1. ICH. (2009). *ICH Harmonised Tripartite Guideline: Pharmaceutical Development Q8(R2)*. Retrieved February 2022 from http://www.ich.org/
2. ICH. (2011). *ICH Harmonised Tripartite Guideline: Development and Manufacture of Drug Substances (Chemical Entities and Biotechnological/Biological Entities) Q11*. Retrieved February 2022 from http://www.ich.org/
3. Shaw, A. D., Kaderbhai, N., Jones, A., Woodward, A. M., Goodacre, R., Rowland, J. J., & Kell, D. B. (1999). Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive Raman spectroscopy and chemometrics. *Applied Spectroscopy*, *53*(11), 1419–1428. https://doi.org/10.1366/0003702991945777
4. Bielser, J. M., Wolf, M., Souquet, J., Broly, H., & Morbidelli, M. (2018). Perfusion mammalian cell culture for recombinant protein manufacturing – a critical review. *Biotechnology Advances*, *36*(4), 1328–1340. https://doi.org/10.1016/j.biotechadv.2018.04.011
5. Vojinović, V., Cabral, J. M. S., & Fonseca, L. P. (2006). Real-time bioprocess monitoring. *Sensors & Actuators, B*, *114*(2), 1083–1091. https://doi.org/10.1016/j.snb.2005.07.059
6. Abu-Absi, N. R., Kenty, B. M., Cuellar, M. E., Borys, M. C., Sakhamuri, S., Strachan, D. J., Hausladen, M. C., & Li, Z. J. (2011). Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. *Biotechnology and Bioengineering*, *108*(5), 1215–1221. https://doi.org/10.1002/bit.23023
7. André, S., Cristau, L. S., Gaillard, S., Devos, O., Calvosa, É., & Duponchel, L. (2015). In-line and real-time prediction of recombinant antibody titer by in situ Raman spectroscopy. *Analytica Chimica Acta*, *892*, 148–152. https://doi.org/10.1016/j.aca.2015.08.050
8. Bhatia, H., Mehdizadeh, H., Drapeau, D., & Yoon, S. (2018). In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models. *Engineering in Life Sciences*, *18*(1), 55–61. https://doi.org/10.1002/elsc.201700084
9. Kozma, B., Salgó, A., & Gergely, S. (2018). Comparison of multivariate data analysis techniques to improve glucose concentration prediction in mammalian cell cultivations by Raman spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, *158*, 269–279. https://doi.org/10.1016/j.jpba.2018.06.005
10. Webster, T. A., Hadley, B. C., Dickson, M., Busa, J. K., Jaques, C., & Mason, C. (2021). Feedback control of two supplemental feeds during fed-batch culture on a platform process using inline Raman models for glucose and phenylalanine concentration. *Bioprocess and Biosystems Engineering*, *44*(1), 127–140. https://doi.org/10.1007/s00449-020-02429-y
11. Whelan, J., Craven, S., & Glennon, B. (2012). In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnology Progress*, *28*(5), 1355–1362. https://doi.org/10.1002/btpr.1590
12. Santos, R. M., Kessler, J. M., Salou, P., Menezes, J. C., & Peinado, A. (2018). 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*(3), 659–670. https://doi.org/10.1002/btpr.2635
13. Berry, B., Moretto, J., Matthews, T., Smelko, J., & Wiltberger, K. (2015). Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. *Biotechnology Progress*, *31*(2), 566–577. https://doi.org/10.1002/btpr.2035
14. Rowland-Jones, R. C., & Jaques, C. (2019). At-line raman spectroscopy and design of experiments for robust monitoring and control of miniature bioreactor cultures. *Biotechnology Progress*, *35*(2), e2740. https://doi.org/10.1002/btpr.2740
15. Rafferty, C., Johnson, K., O'Mahony, J., Burgoyne, B., Rea, R., & Balss, K. M. (2020). Analysis of chemometric models applied to Raman spectroscopy for monitoring key metabolites of cell culture. *Biotechnology Progress*, *36*(4), 1–16. https://doi.org/10.1002/btpr.2977
16. Rowland-Jones, R. C., Graf, A., Woodhams, A., Diaz-Fernandez, P., Warr, S., Soeldner, R., Finka, G., & Hoehse, M. (2021). Spectroscopy integration to miniature bioreactors and large scale production bioreactors – increasing current capabilities and model transfer. *Biotechnology Progress*, *37*(1), 1–14. https://doi.org/10.1002/btpr.3074
17. Domján, J., Fricska, A., Madarász, L., Gyürkés, M., Köte, Á., Farkas, A., Vass, P., Fehér, C., Horváth, B., Könczöl, K., Pataki, H., Nagy, Z. K., Marosi, G. J., & Hirsch, E. (2020). Raman-based dynamic feeding strategies using real-time glucose concentration monitoring system during adalimumab producing CHO cell cultivation. *Biotechnology Progress*, *36*(6), e3052. https://doi.org/10.1002/btpr.3052
18. Yilmaz, D., Mehdizadeh, H., Navarro, D., Shehzad, A., O'Connor, M., & McCormick, P. (2020). Application of Raman spectroscopy in monoclonal antibody producing continuous systems for downstream process intensification. *Biotechnology Progress*, *36*(3), e2947. https://doi.org/10.1002/btpr.2947
19. Feidl, F., Garbellini, S., Vogg, S., Sokolov, M., Souquet, J., Broly, H., Butté, A., & Morbidelli, M. (2019). A new flow cell and chemometric protocol for implementing in-line Raman spectroscopy in chromatography. *Biotechnology Progress*, *35*(5), 1–10. https://doi.org/10.1002/btpr.2847
20. Batra, J., & Rathore, A. S. (2016). Glycosylation of monoclonal antibody products: Current status and future prospects. *Biotechnology Progress*, *32*(5), 1091–1102. https://doi.org/10.1002/btpr.2366
21. Eon-Duval, A., Broly, H., & Gleixner, R. (2012). Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. *Biotechnology Progress*, *28*(3), 608–622. https://doi.org/10.1002/btpr.1548
22. Brühlmann, D., Muhr, A., Parker, R., Vuillemin, T., Bucsella, B., Kalman, F., Torre, S., La Neve, F., Lembo, A., Haas, T., Sauer, M., Souquet, J., Broly, H., Hemberger, J., & Jordan, M. (2017). Cell culture media supplemented with raffinose reproducibly enhances high mannose glycan formation. *Journal of Biotechnology*, *252*, 32–42. https://doi.org/10.1016/j.jbiotec.2017.04.026
23. Gramer, M. J., Eckblad, J. J., Donahue, R., Brown, J., Shultz, C., Vickerman, K., Priem, P., van den Bremer, E. T. J., Gerritsen, J., & van Berkel, P. H. C. (2011). Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose. *Biotechnology and Bioengineering*, *108*(7), 1591–1602. https://doi.org/10.1002/bit.23075
24. Surve, T., & Gadgil, M. (2015). Manganese increases high mannose glycoform on monoclonal antibody expressed in CHO when glucose is absent or limiting: Implications for use of alternate sugars. *Biotechnology Progress*, *31*(2), 460–467. https://doi.org/10.1002/btpr.2029
25. Villiger, T. K., Steinhoff, R. F., Ivarsson, M., Solacroup, T., Stettler, M., Broly, H., Krismer, J., Pabst, M., Zenobi, R., Morbidelli, M., & Soos, M. (2016). High-throughput profiling of nucleotides and nucleotide sugars to evaluate their impact on antibody N-glycosylation. *Journal of Biotechnology*, *229*, 3–12. https://doi.org/10.1016/j.jbiotec.2016.04.039

26. Zhang, L., Castan, A., Stevenson, J., Chatzissavidou, N., Vilaplana, F., & Chotteau, V. (2019). Combined effects of glycosylation precursors and lactate on the glycoprofile of IgG produced by CHO cells. *Journal of Biotechnology*, *289*, 71–79. https://doi.org/10.1016/j.jbiotec.2018.11.004
27. Zupke, C., Brady, L. J., Slade, P. G., Clark, P., Caspary, R. G., Livingston, B., Taylor, L., Bigham, K., Morris, A. E., & Bailey, R. W. (2015). Real-time product attribute control to manufacture antibodies with defined N-linked glycan levels. *Biotechnology Progress*, *31*(5), 1433–1441. https://doi.org/10.1002/btpr.2136
28. Metze, S., Ruhl, S., Greller, G., Grimm, C., & Scholz, J. (2020). Monitoring online biomass with a capacitance sensor during scale-up of industrially relevant CHO cell culture fed-batch processes in single-use bioreactors. *Bioprocess and Biosystems Engineering*, *43*(2), 193–205. https://doi.org/10.1007/s00449-019-02216-4
29. Lai, B., Plan, M. R., Hodson, M. P., & Krömer, J. O. (2016). Simultaneous determination of sugars, carboxylates, alcohols and aldehydes from fermentations by high performance liquid chromatography. *Fermentation*, *2*(1), 6. https://doi.org/10.3390/fermentation2010006
30. Dudek, M., Zajac, G., Szafraniec, E., Wiercigroch, E., Tott, S., Malek, K., Kaczor, A., & Baranska, M. (2019). Raman optical activity and Raman spectroscopy of carbohydrates in solution. *Spectrochimica Acta – Part A: Molecular and Biomolecular Spectroscopy*, *206*, 597–612. https://doi.org/10.1016/j.saa.2018.08.017
31. Dumouilla, V., & Dussap, C. G. (2021). Online analysis of D-glucose and D-mannose aqueous mixtures using Raman spectroscopy: An in silico and experimental approach. *Bioengineered*, *12*(1), 4420–4431. https://doi.org/10.1080/21655979.2021.1955550

SUPPORTING INFORMATION

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

How to cite this article: Romann, P., Kolar, J., Tobler, D., Herwig, C., Bielser, J.-M., & Villiger, T. K. (2022). Advancing Raman model calibration for perfusion bioprocesses using spiked harvest libraries. *Biotechnology Journal*, e2200184. https://doi.org/10.1002/biot.202200184