Research Article

Received: 2 September 2019 Revised: 24 November 2019 Accepted article published: 3 December 2019 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jctb.6299

Osmolality is a predictor for model-based real time monitoring of concentration in protein chromatography

Edith Felfödi,[a] Theresa Scharl,[a,b] Michael Melcher,[a,b] Astrid Dürauer,[a,c] Kristeena Wright[d] and Alois Jungbauer[a,c*]

Abstract

BACKGROUND: The bottleneck for real time control and real time release is lack of product-specific in-line sensors or fast at-line methods suitable for model-based prediction of process outcome. The most common sensors for protein purification are UV absorbance values measured at 280 and 260 nm. They have very high selectivity for proteins which contain aromatic amino acids. The 260 nm signal is more selective for nucleic acids. This work addresses the question if osmolality can be used as an additional predictor for protein purification.

RESULTS: An antibody intermediate purification step in flow-through mode was evaluated. The flow-through fractions were collected and then subjected to analysis for antibody concentration and osmolality. UV280, UV260, UV214, pH and conductivity have been measured on-line by the chromatography workstation. Different combinations of on-line sensor signals and osmolality have been used to find out if molality is a valuable predictor. The root mean square error was used for assessing the quality of the model-based prediction of quantity with partial least squares in this chromatography process. Predictors UV280, UV260, UV214, pH and conductivity showed equal root mean square error (0.274) as UV280, UV260, UV214, pH, conductivity and osmolality (0.274). Lowest mean square error (0.244) was found with UV280, UV260 and osmolality as predictors of quantity.

CONCLUSION: Osmolality as an at-line method is an excellent predictor together with UV280 and UV260 for protein quantity in model-based prediction using partial least squares methodology.

Keywords: predictor; osmolality; freezing point depression; antibody; purification; partial least squares

INTRODUCTION

Real time monitoring and real time release of protein purification are encouraged by health authorities. Guidelines of the European Health Agency and the US Food and Drug Administration have been in place for several years.[1] 'Real time release testing' is the ability to evaluate and ensure the quality of in-process and/or final product based on process data. Typically, it includes a valid combination of measured material attributes and process controls. The rules are outlined in the guidelines of the International Conference of Harmonization (ICH Q8 (R2)).[2] Real time release testing improves product quality and consistency.[3,4] The bottleneck for real time control and real time release is the lack of product-specific in-line sensors or fast at-line methods. In this work, we want to check if osmolality is a valuable predictor for protein purification. One possibility to circumvent the lack of a product-specific sensor is the application of statistical modeling, often referred to as soft sensors.[5] On-line and off-line measurements in a process are correlated by statistical models, and the desired properties such as product concentration are then predicted. This requires the training of the model, usually done by several process runs. This has been demonstrated for purification of model proteins [6–8] or real process samples such as antibodies from culture supernatant by protein A affinity chromatography[9,10] or basic fibroblast growth factor by cation exchange chromatography.[11] The most common sensors for protein purification are UV absorbance values measured at 280 and 260 nm. They have very high selectivity for proteins which contain aromatic amino acids. The 260 nm signal is more selective for nucleic acids.[12] pH and conductivity are not able to directly measure protein concentration in protein

* *Correspondence to: A Jungbauer, Institute of Bioprocess Science and Engineering, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria. E-mail: alois.jungbauer@boku.ac.at*

a *Austrian Centre of Industrial Biotechnology, Vienna, Austria*

b *Institute of Statistics, University of Natural Resources and Life Sciences Vienna, Vienna, Austria*

c *Institute of Bioprocess Science and Engineering, University of Natural Resources and Life Sciences Vienna, Vienna, Austria*

d *Advanced Instruments, Norwood, A, USA*

chromatography, but with change of salt and/or pH, proteins are eluted from the column. They are indirect predictors.

We raised the question as to whether osmolality could be an additional predictor for protein purification. The osmolality of a solution denotes the concentration of osmotically active particles in that solution.[13] Osmolality is defined as the concentration of all solutes in a given mass of water and is expressed as units of either osmolality (milliosmoles of solute per kilogram of water, mOsm kg^{-1} H_2O) or osmolarity (milliosmoles of solute per liter of water, mOsm L^{-1} H_2O):

$$\text{Osmolality} = \phi nC = \text{osmol/kg } H_2O \quad (1)$$

where ϕ is the osmotic coefficient, which accounts for the degree of molecular dissociation, n is the number of particles into which a particle can dissociate and C is the molal concentration of the solution. Osmolality is a function only of the number of particles and is not related to the particles' molecular mass, size, shape or charge. One mole of a non-dissociating substance (e.g. glucose or urea) dissolved in 1 kg of water decreases the freezing point of the resultant solution by 1.86 °C. Such a solution has an osmolality of 1 Osm kg^{-1} or 1000 mOsm kg^{-1}. The presence of solutes in an aqueous solution changes the colligative properties: increase of boiling point, decrease of freezing protein, increase of osmotic pressure and decrease of vapor pressure. Osmolality is conveniently measured by freezing point osmometry, vapor pressure osmometry or membrane osmometry. We have applied freezing point technology because it is the industry-expected osmolality test method (EP 2.2.35: Osmolality) and offers fast test times as well as accuracy and precision.

Our goal was to enable the prediction of product concentration in real time. We only used a small set of predictors that are typically recorded in a time grid of 1 s by the chromatography workstation, i.e. three UV signals (UV280, UV260 and UV214), pH and conductivity, together with osmolality. As previously shown,[9] partial least squares (PLS) regression models are able to predict the quantity and purity of an antibody capture step. In order to evaluate if osmolality is a useful predictor for protein concentration, several PLS prediction models with and without osmolality were generated. As we currently only had eight antibody purification runs available where osmolality was measured, the prediction models were optimized using leave-one-run-out cross-validation. Seven runs were used as a training set and then the models were applied to the eighth run. The performance of the different prediction models was compared using the root mean square error of prediction (RMSE). The RMSE is given in the unit of the variable of interest and measures the average deviation between measured and predicted protein concentration for all fractions. When designing a dataset for the prediction of protein concentration, it is important to find a tradeoff between the size and the number of the fractions used. The fractions should be large enough to be able to perform all necessary analytical analysis and small enough to capture the shape of the elution. In addition, the measurement error of the analytical analysis has to be considered.

In this work, we used a flow-through chromatography step to determine if osmolality could be an additional predictor for protein concentration. We have set up a flow-through intermediate step for antibody purification as a model process using a mixed mode anion exchanger. The eluate from protein A affinity chromatography is diluted and adjusted to pH 8.75 and loaded in the mixed mode ion exchanger. The flow-through fractions are collected and then subjected to analysis for antibody concentration and osmolality. UV280, UV260, UV214, pH and conductivity have been measured on-line by the chromatography workstation. Different combinations of on-line sensor signals and osmolality have been used to find out by statistical modeling if osmolality is a valuable predictor.

EXPERIMENTAL

Chromatographic workstation

The chromatographic workstation consisted of an ÄKTA Pure 25 system (GE Healthcare, Uppsala, Sweden) equipped with two system pumps, sample pump S9 and fraction collector F9-C. It also contained the standard UV monitor U9-M for triple-wavelength detection as well as the conductivity probe C9 and the pH probe V9-pH. The system was controlled by the software Unicorn Version 6.4. UV absorption at 280, 260 and 214 nm was recorded using a flow cell with an optical path length of 2 mm.

Antibody and capture step

A monoclonal antibody against tumor necrosis factor alpha was produced in CHO K1 cell culture. The antibody was captured from the supernatant via protein A affinity chromatography using an antibody Select SuRe column (GE Healthcare) with glycine/HCl (pH 3.5) as elution buffer. The fractions were neutralized by adding 1/20 fraction volume of 0.5 mol L^{-1} sodium phosphate (pH 8.0) and pooled. The pool was aliquoted, then the 5 mL aliquots were frozen and stored at −20 °C until further use. The antibody concentration of the aliquoted antibody pool was 12.1 mg mL^{-1}.

Antibody intermediate purification using TOYOPEARL NH2-750F mixed mode resin

TOYOPEARL NH2-750F (Tosoh Bioscience, Griesheim, Germany) is a salt-tolerant anion exchanger resin; however, as the ligand is a polyamine chain, it is considered to have mixed mode interactions. The resin was packed into a Tricorn column with 5 mm inner diameter to a total volume of 1 mL. The intermediate step was operated in flow-through mode, retaining host cell proteins and high-molecular-weight impurities. Tris/HCl (10 mmol L^{-1}, pH 8.75) was used as equilibration buffer, 10 mmol L^{-1} Tris/HCl containing 1 mol L^{-1} NaCl (pH 8.75) served as regeneration buffer and 0.1 mol L^{-1} NaOH was used for sanitization of the column.

On the day of usage, the 5 mL antibody capture eluate was thawed and diluted 1:5 with equilibration buffer. The pH was adjusted to 8.75 by adding few drops of 10 mol L^{-1} NaOH. The load was filtered using a syringe filter with 0.2 µm pore size prior to loading onto the column. After the 1:5 dilution with equilibration buffer, the load had an antibody concentration of approximately 2.4 mg mL^{-1}. During sample loading, 1.2 mL flow-through fractions were collected. The flow rate was 100 cm h^{-1}.

Osmolality measurement

The osmolality of each flow-through fraction was measured using an OsmoTECH® Single-Sample Micro-Osmometer (Advanced Instruments, Norwood, MA, USA). The technique of measurement was based on the determination of freezing point depression. For the measurement, 20 µL of the sample was aspirated and the sampler was inserted directly into the instrument. Samples were run in duplicate and osmolality values were recorded and analyzed.

Journal of Chemical Technology & Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

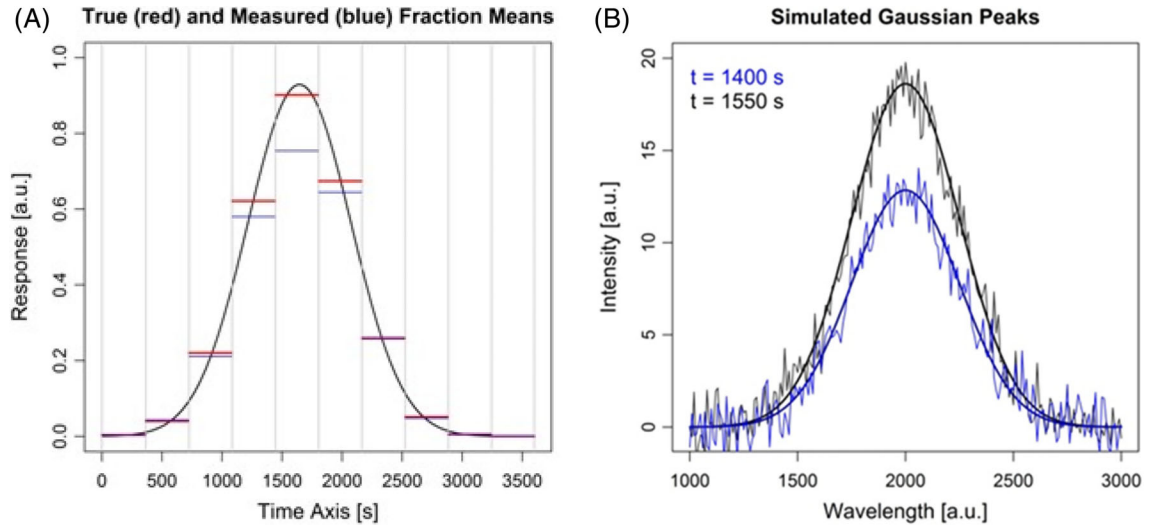

Figure 1. (A) Theoretical protein elution profile (black), corresponding fraction-wise averages (red) for $n_f = 10$ fractions and actual measured responses (blue). (B) Noisy simulated spectra (given for two time points) serve as predictors (a.u. is arbitrary units).

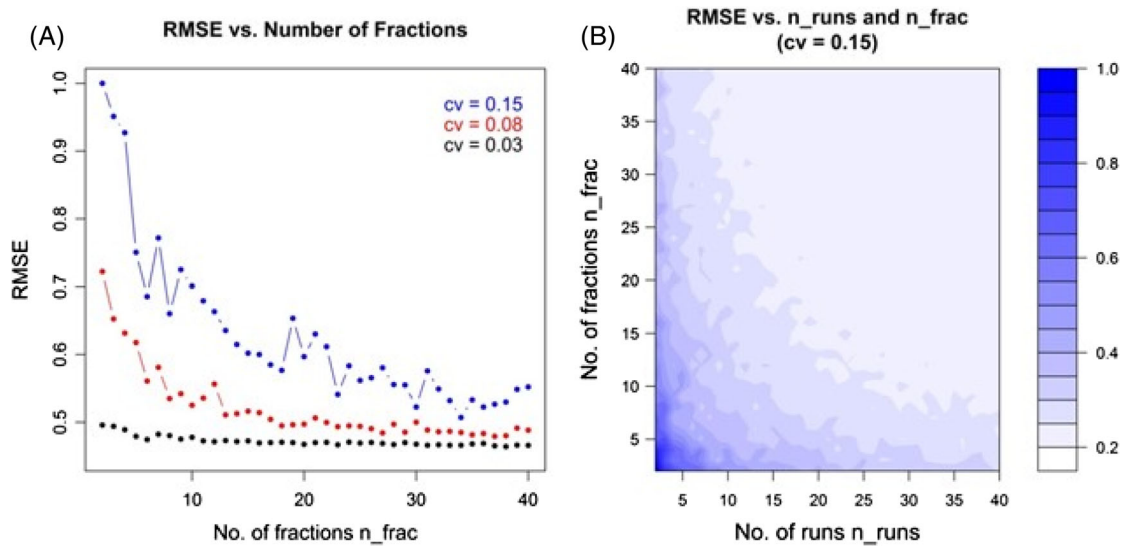

Figure 2. (A) Modeling error as a function of number of fractions per run (n_f) for values of coefficient of variation (*cv*) representing accuracy of analytical method. (B) Pseudo-3D contour plot with prediction error as a function of both n_r and n_f.

Statistical model

The eight identical repetitions of the antibody intermediate process were used to set up a statistical prediction model. The on-line data used were recorded in a time grid of 1 s by the chromatography workstation, i.e. three UV signals (UV280, UV260 and UV214), pH and conductivity. The on-line data were stored in the database of the custom programmed control software XAMIris (Evon, Austria). The data were read into the statistical computing environment R where all statistical analyses were performed. First, quality and consistency checks of the data were performed by visual inspection together with descriptive statistics. For model building, the time intervals of the on-line data were reduced by averaging across the time grid of the off-line analytics (antibody content, aggregates and monomers) and at-line measurement of osmolality. In addition, the time delay by the void volumes of the pH and conductivity sensor was considered.

PLS regression is a well-established modeling technique in chemometrics. PLS transforms the original predictor variables into a set of latent variables, which are also linear functions of the original predictors. These latent variables are then used as predictors in a multiple linear regression model for the response *y*. Linear combinations are determined such that a maximum covariance between the scores (the values of the latent variables) and the response is achieved. The number of components (latent variables) is an optimization parameter and was determined within the framework of a leave-one-run-out cross-validation – i.e. one of the eight runs is held out and predicted by a model fit on the remaining seven runs. This procedure is repeated until each run has served as a test run once. The output of this algorithm is the optimized model parameter and a prediction error estimate. The RMSE was calculated according to the following equation:

$$\mathrm{RMSE} = \left[\frac{1}{k}\sum_{j=1}^{k}(y_i - \hat{y}_i)^2\right]^{1/2} \quad (2)$$

Journal of Chemical Technology & Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

where k is the number of observations, y_i represents the measured values and $\hat{y}_i$ represents the predicted values.

For PLS, scaling is indispensable as all the different variables were measured on different scales and units. In this study, autoscaling was performed, where each variable x was transformed individually by subtraction of the mean of x and division by the empirical standard deviation.

RESULTS AND DISCUSSION

The development of a statistical model for prediction of a process parameter from on-line and at-line analysis requires training runs. This means that a process is repeated several times and the on-line, at-line and off-line analysis is performed. With this dataset then, a statistical model for prediction of a process parameter is established. In this case, a flow-through chromatography purification of an antibody was performed with an ÄKTA chromatography workstation. This workstation was equipped with UV sensors (280, 260 and 214 nm), a conductivity monitor and a pH probe. At-line, the osmolality was measured. The off-line analytics were antibody content, aggregates and monomers.

First we determined the optimal number of fractions for the off-line analysis, because the flow-through is divided into fractions. Therefore a theoretical assessment was done by assuming decreasing accuracy expressed as coefficient of variation (*cv*) of our off-line methods. We use *cv* as fraction of one and not %.

For the training of the statistical model, it is also important to select the appropriate number of runs. The more precise the off-line methods are, the better the quality of prediction will be. All analytical methods perform with a certain precision. Therefore a compromise must be found between the number of runs, the number of collected fractions and the expected quality of prediction expressed as RMSE. More training runs will improve the quality of prediction for a given precision of an analytical method. A compromise must be made in order to keep the number of experiments to a reasonable number. A *cv* of 0.03 would represent a method with high precision, such as high-performance liquid chromatography (HPLC),[14,15] and a *cv* of 0.15 would represent a method such as enzyme-linked immunosorbent assay (ELISA).

Statistical modeling requires data – predictors x_i (usually p-vectors with entries being the UV intensities, pH, conductivity or osmolality) and corresponding measurements of the response y_i (e.g. the protein quantity) for a set of k observations originating from n_r chromatographic runs forming the so-called training set. In the present case, the predictors are available at a high frequency of typically a few seconds, whereas the protein quantity can only be determined fraction-wise (i.e. in a cumulative manner). An obvious solution to this problem is the averaging of the on-line signals over the time interval of the off-line fraction to obtain the required matching (x_i, y_i) data pairs. Natural questions in this context are the number of runs, n_r, and the number of fractions per run, n_f, to be performed to achieve a low prediction error at an acceptable workload in the laboratory and determined by the product $k = n_r n_f$. In a simulation experiment, we assume a Gaussian profile of the protein quantity (the 'true' values) as shown in Fig. 1A, the corresponding means (red), which would be determined in the laboratory in the absence of any experimental error, and the actually measured values (blue). The latter are simulated for a given *cv* assuming an unbiased analytical method (i.e. its long-run average is identical to the true value). Similarly, noisy artificial and unimodal spectra are simulated on a 5 s time grid (Fig. 1B) with a linear relationship between the peak height and the protein quantity.

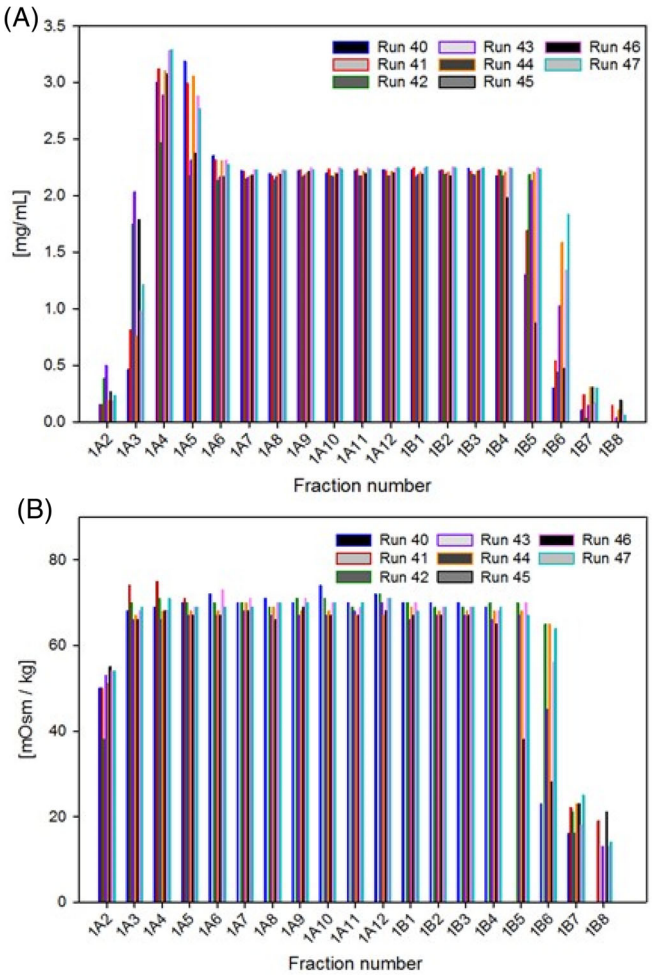

Figure 3. (A) Antibody concentration in eight training runs. 1A2–1B8 are the collected fractions. (B) Osmolality of different fractions.

PLS models are built on a training set given by n_r runs, n_f fractions per run and a measurement error given by an assumed *cv*, optimized for the number of PLS components by 10-fold cross-validation, and the final model applied on a (large and simulated) test set. As both the absolute values of the response and the noise level in the predictor spectra are arbitrary, any achieved results are only relative. Figure 2A depicts the test error (RMSE) as a function of the number of fractions for different values of *cv* and a fixed number of runs. As expected, the modeling error decreases with n_f and is generally lower for more accurate analytical methods. As a first conclusion, a flattening of all profiles for n_f larger than 10–20 can be seen, with the exact value depending on *cv*. Secondly, Fig. 2B shows that lines with a constant modeling error have an approximate hyperbolic shape of the form $n_f n_r$ = constant. The practical implication of this finding is as follows. One might start modeling with a dataset obtained for a certain value of n_f and observe the resulting RMSE. If this seems too high, performing more runs (i.e. increasing n_r) will improve the model performance.

With approximately eight training runs and about 18 fractions, we are already in a good range of quality of prediction. Once we had done this theoretical assessment, we performed the training runs. With the eight training runs, we measured the antibody

Journal of Chemical Technology & Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

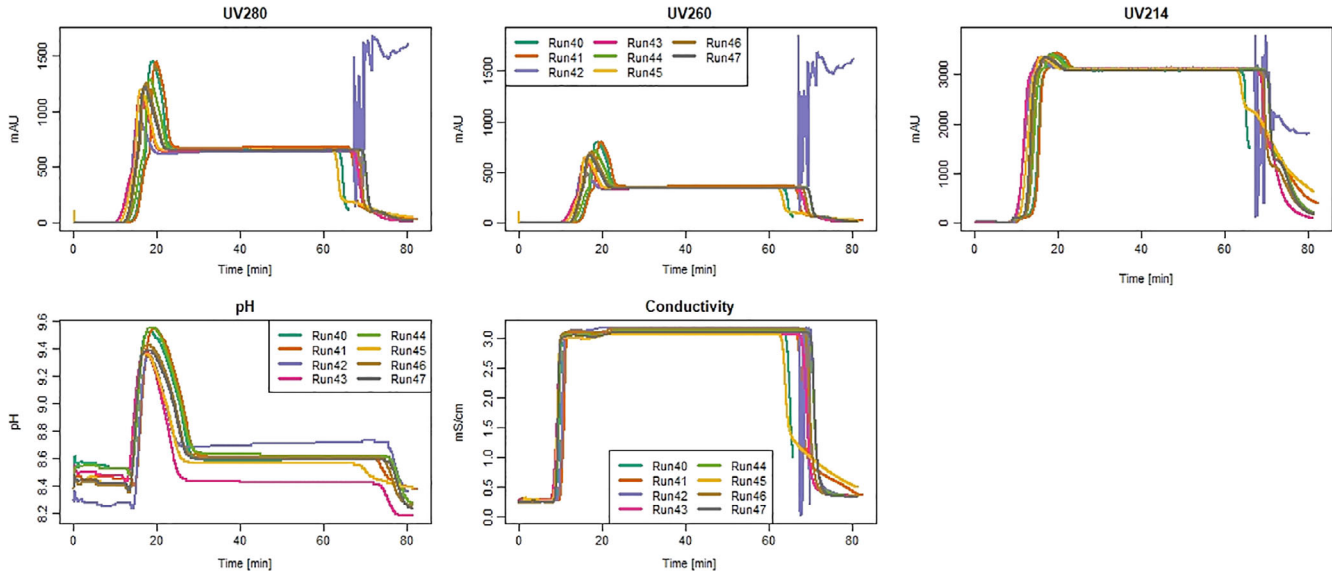

Figure 4. On-line UV signals, pH signal and conductivity profile for eight training runs.

concentration in the collected fractions by analytical affinity chromatography (Fig. 3A). The training runs should have the same accuracy as later on the runs for production.

The roll-up in the third and fourth fractions is explained by displacement of the antibody monomers by antibody multimers which bind stronger than the monomers or by the pH transition which is generated upon elution. Fractions were also analyzed for osmolality (Fig. 3B). These measurements were treated as at-line measurements. Unfortunately, osmolality is not currently available as an on-line sensor, therefore the test must be processed at-line in order to be able to assess if osmolality is a predictor for protein concentration in flow-through chromatography.

On-line signals UV280, UV260, UV214, pH and conductivity are shown in Fig. 4. The signal traces reflect the day-to-day variation of an antibody intermediate purification and are therefore well suited for model building.

With these training runs, the statistical model was established to predict quantity for this purification process. Prediction models were generated with various combinations of predictors. The number of latent variables was optimized using leave-one-run-out cross-validation. In the six presented models, the cross-validated results, i.e. for the prediction of one single run the data of the seven remaining runs, were used for the training of the model. The predicted and measured data were superimposed.

The quantity can be predicted with good accuracy by the predictors UV280, UV260, UV214, pH and conductivity (Fig. 5, PM1). RMSE is given in units of quantity (mg mL^{-1}), and here it is less than 10% of the maximum antibody concentration. Addition of osmolality to the predictors did not improve the quality of prediction. The same RSME was obtained (Fig. 5, PM2). The combination of UV280, UV260 and osmolality (Fig. 5, PM3) or UV280, UV260, UV214 and osmolality (Fig. 5, PM4) improved the quality of prediction. It is obvious that osmolality alone failed to predict quantity (Fig. 5, PM5). However, it is slightly better than conductivity alone (Fig. 5, PM6). UV is also a good predictor for quantity, because it captures the aromatic amino acids of proteins and the SH groups. It is often overlooked that the SH groups also contribute to UV absorbance. UV and pH increase at the beginning of the flow-through. Overshoot/roll-up is very common in flow-through chromatography, because two different protein species compete with each other, resulting in the weaker binding species being displaced by the stronger one. When the protein is applied to the equilibrated ion exchanger, salt and hydroxide ions are displaced by the protein which is weakly bound. This leads to an increase in pH and a slight increase in conductivity. The pH monitor is slow compared with UV. This explains why the combinations of UV280, UV260 and osmolality (Fig. 5, PM3) and UV280, UV260, UV214 and osmolality (Fig. 5, PM4) are slightly better predictors. The slow pH response will predict a later onset of protein elution and a broader roll-up. Furthermore, we worked with weak buffers that were not fully dissociated. A slight change in pH has a dramatic effect on dissociation. The protein itself acts as a buffer. These effects are captured by osmolality. Although the effect of osmolality is well documented for cell culture, it has been neglected in protein chromatography. Currently, osmometers are batch instruments, and on-line sensors are not yet available. On the other hand, an at-line measurement may contribute to better process understanding, because nowadays measurements with modern osmometers are not very time-consuming and only a few microliters are necessary for a precise measurement. The method captures all solutes in a chromatography process solution, the ions dissociated and non-dissociated, proteins and non-ionic solutes such as sugars and polyols. pH captures only hydronium and hydroxide ions, and conductivity only charged ions.

CONCLUSION

Osmolality as an at-line method is an excellent predictor together with UV280 and UV260 for protein quantity in model-based prediction using PLS methodology. At the current stage, it can be used as an at-line method because the instrument gives a very fast response. Osmolality is independent of temperature because it measures osmolytes based on mass, whereas pH and conductivity are highly dependent on temperature. All these considerations explain why osmolality is a slightly better predictor together with UV compared with UV, pH and conductivity.

Journal of Chemical Technology & Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

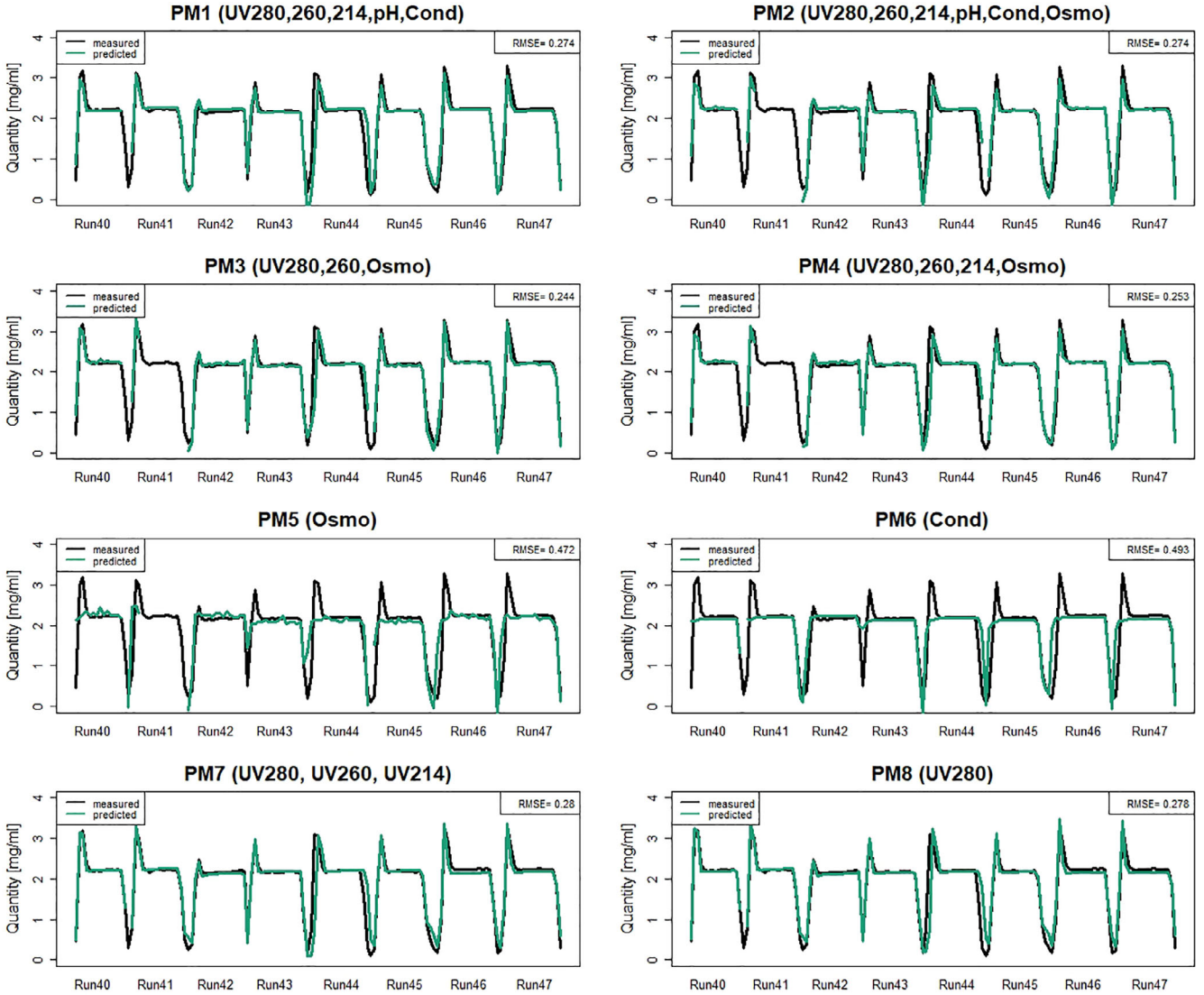

Figure 5. Prediction of antibody quantity by various predictor sets.

ACKNOWLEDGEMENTS

E.F., M.M. and T.S. have been supported in part by the Federal Ministry for Digital and Economic Affairs (bmwd), the Federal Ministry for Transport, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, Government of Lower Austria and ZIT – Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. The funding agencies had no influence on the conduct of this research.

REFERENCES

1 EMA, *Guideline on Real Time Release Testing (formerly Guideline on Parametric Release)*. European Medicines Agency, London (2012).
2 ICH, *ICH Harmonized Tripartite Guideline: Pharmaceutical Development, Q8 (R2)*. ICH, Geneva (2009).
3 Jiang M, Severson KA, Love JC, Madden H, Swann P, Zang L *et al.*, Opportunities and challenges of real-time release testing in biopharmaceutical manufacturing. *Biotechnol Bioeng* **114**:2445–2456 (2017).
4 Guerra A, von Stosch M and Glassey J, Toward biotherapeutic product real-time quality monitoring. *Crit Rev Biotechnol* **39**:289–305 (2019).
5 Roch P and Mandenius C-F, On-line monitoring of downstream bioprocesses. *Curr Opin Chem Eng* **14**:112–120 (2016).
6 Brestrich N, Briskot T, Osberghaus A and Hubbuch J, A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. *Biotechnol Bioeng* **111**:1365–1373 (2014).
7 Brestrich N, Sanden A, Kraft A, McCann K, Bertolini J and Hubbuch J, Advances in inline quantification of co-eluting proteins in chromatography: process-data-based model calibration and application towards real-life separation issues. *Biotechnol Bioeng* **112**:1406–1416 (2015).
8 Großhans S, Rüdt M, Sanden A, Brestrich N, Morgenstern J, Heissler S *et al.*, In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography. *J Chromatogr A* **1547**:37–44 (2018).
9 Walch N, Scharl T, Felföldi E, Sauer DG, Melcher M, Leisch F *et al.*, Prediction of the quantity and purity of an antibody capture process in real time. *Biotechnol J*:14, 7:1800521 (2019).
10 Rüdt M, Brestrich N, Rolinger L and Hubbuch J, Real-time monitoring and control of the load phase of a protein A capture step. *Biotechnol Bioeng* **114**:368–373 (2017).
11 Sauer DG, Melcher M, Mosor M, Walch N, Berkemeyer M, Scharl-Hirsch T *et al.*, Real-time monitoring and model-based prediction of purity and quantity during a chromatographic capture of fibroblast growth factor 2. *Biotechnol Bioeng* **116**:1999–2009 (2019).

12 Rüdt M, Briskot T and Hubbuch J, Advances in downstream processing of biologics – spectroscopy: an emerging process analytical technology. *J Chromatogr A* **1490**:2–9 (2017).
13 Johnson CW, Timmons DL and Hall PE, *Essential Laboratory Mathematics: Concepts and Applications for the Chemical and Clinical Laboratory Technician*, 2nd edn. Delmar Cengage Learning, Clifton Park, NY (2003).
14 Satzer P and Jungbauer A, High-capacity protein A affinity chromatography for the fast quantification of antibodies: two-wavelength detection expands linear range. *J Sep Sci* **41**:1791–1797 (2018).
15 Tscheliessnig A and Jungbauer A, High-performance monolith affinity chromatography for fast quantitation of immunoglobulin G. *J Chromatogr A* **1216**:2676–2682 (2009).

Journal of Chemical Technology & Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.