Journal of Chromatography A, 1257 (2012) 41–47

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Rapid quantification of protein–polyethylene glycol conjugates by multivariate evaluation of chromatographic data

Sigrid K. Hansen, Benjamin Maiser, Jürgen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 19 April 2012
Received in revised form 23 July 2012
Accepted 24 July 2012
Available online 3 August 2012

Keywords:
Multivariate calibration
High throughput analytics
High throughput experimentation
PEGamer quantification
Protein PEGylation
High throughput process development

ABSTRACT

Size exclusion chromatography (SEC) is often applied for characterization of protein–polyethylene glycol (PEG) conjugates regarding the number of attached PEG chains (PEGamers). SEC analysis is advantageous as it is precise, robust, and straightforward to establish. However, most SEC based assays have a maximal throughput of a few samples per hour. We present a strategy to increase analytical throughput based on combining a short column with a fast flow rate, and finally multivariate calibration in order to compensate for the resolution lost in the trade off for speed. Different multivariate approaches were compared and multilinear regression was shown to result in the most precise calibrations. Further, a dynamic calibration approach was developed in order to account for changes in column performance over time. In this way, it was possible to establish a highly precise assay for protein PEGamer quantification with a throughput of 30 samples per hour.

1. Introduction

Attaching polyethylene glycol (PEG) polymer chains to proteins (PEGylation) has been shown to improve their pharmacokinetics. The increase in size reduces renal clearance of the therapeutic and there is proof of immunogenicity and antigenicity being reduced by PEGylation [1,2]. Further, solubility of hydrophobic proteins can be increased by PEGylation [3]. The PEGylation process determines the PEGamer distribution and positions of the PEG molecules attached to the protein. If random PEGylation is performed, the product of the reaction will be very heterogeneous. This poses a problem in a regulative environment which demands defined products of extremely high homogeneity. In general, there are two ways to reach a homogeneously PEGylated product. Either the product mixture of the random PEGylation can be purified to contain only the desired PEGylated species or site specific PEGylation can be performed.

In our lab, small scale PEGylation performed automated on liquid handling stations has been established. This enables high throughput screenings of PEGylation process parameters and is a part of a general trend where automated high throughput experimentation (HTE) is applied for high throughput process development (HTPD) of biologicals [4]. For the evaluation of the performed experiments to stand in relation to experimental speed when performing HTE, the analytical throughput must match the experimental throughput. In some cases this comes easily, for instance if the evaluation of screenings performed in HTE mode can be based merely on univariate spectroscopic measurements such as total protein quantification via UV absorption measurements [5–7]. If selective or specific quantification is necessary, other methods have to be considered. In the case of quantitative separation of protein PEGamers, size exclusion chromatography (SEC) is widely applied. However, for SEC based assays to match the speed of HTE, the analysis time per sample must be reduced to merely a few minutes.

SEC is a standard method for selective and specific quantification of proteins. If exact quantification is the primary objective, most analysts will seek to achieve high resolution ($R > 1.5$) of the components. If a short analysis time is the primary objective, faster flow rates and/or shorter columns can be applied, however, at the expense of resolution. One approach to increase analytical speed without losing resolution is the interlacing of injections and/or parallel operation of two columns on a chromatographic system [8,9]. When interlaced injections are performed in SEC analysis, the result is elimination of initial lag time between sample injection and start of elution. If interlaced injections are combined with parallel operation of two columns, the waiting time post elution of the smallest sample molecule of interest can be reduced or even eliminated. Thus, in favorable situations, analysis time per sample can be reduced to the time span in which the molecules of interest elute.

The separation of protein PEGamers using SEC poses a very challenging task, as the relative increase in molecular size decreases for each additional PEG molecule attached to the protein [10].

* Corresponding author. Tel.: +49 721 608 42557; fax: +49 721 608 46240.
E-mail address: juergen.hubbuch@kit.edu (J. Hubbuch).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.07.089

A size based chromatographic separation of native protein and mono-PEGylated species might be achieved effortlessly, however a resolution of the higher protein PEGamers will become increasingly difficult.

Our proposition is that for analytical purposes the degree of separation necessary for correct quantification can be decreased significantly by applying alternative methods for the evaluation of chromatographic data. This idea is based on the assumption that overlapping elution of different species will result in a chromatogram which is a linear superposition of the signals of each single analyte. Therefore, a linear multivariate correlation between elution profile (chromatogram) and amounts of the different components in the respective samples can be expected. This, of course, presumes that the amount of the analytes does not influence the elution profile. In the linear range of adsorption, this assumption would apply as peak shape is a function of the chromatography system (dead volume, column packing *etc.*) and protein characteristics (adsorption and diffusion) but not of protein concentration. Further, the separation mechanism of SEC does not depend on direct interaction with the column material and thus it is not expected that different load concentrations should cause a non-linear change of the elution profile. If the linear correlation between analyte concentration and elution profile is given, it should be possible to calibrate a multivariate regression model based on one chromatogram of each pure component and one defined mixture of all components. Such a multivariate regression model could then give precise determinations of sample composition despite low chromatographic resolution of the sample components.

The presented work describes how multivariate calibration can be used to gain quantitative results of high quality from low resolution chromatograms. By doing so, faster analysis times are achievable through application of short columns and high flow rates that would otherwise be avoided due to inferior separation performance.

In general, a balanced ratio of analytical and experimental time allows a more efficient use of the HTE platform. The goal was to achieve an assay time of maximum 2 min. This would facilitate a 24 h experiment-analysis cycle for PEGylation screenings in 96 well HTE format, where screening experiments which are performed during day time can be analyzed over night. Further, fast assays can be useful for PEGylation reaction monitoring and subsequent purification process monitoring.

2. Materials and methods

2.1. Chemicals

Potassium phosphate, potassium chloride and analytical grade ethanol for SEC buffer preparation were purchased from Merck (Darmstadt, Germany). Lysozyme was purchased from Sigma–Aldrich (St. Louis, MO, USA). Methoxy-PEG aldehyde with an average molecular weight of 5 kDa was provided from NOF Cooperation (Tokyo, Japan). Sodium phosphate, sodium chloride and sodium cyanoborohydride ($NaCNBH_3$) for PEGylation buffer and IEC buffer preparation were purchased from Merck (Darmstadt, Germany).

2.2. PEGylated lysozyme

2.2.1. PEGylation reaction

Lysozyme (5 g/L) and PEG were dissolved in a 25 mM sodium phosphate buffer pH 7.2, containing 20 mM sodium cyanoborohydride. The molar polymer to protein ratio was set to 6:1. The reaction was carried out in a continuously shaken falcon tube at room temperature, for 10 h.

2.2.2. Preparative separation of lysozyme PEGamers

Single lysozyme PEGamers were purified using cation exchange chromatography. Toyopearl GigaCap S-650M resin (Tosoh Biosience GmbH, Stuttgart, Germany) was packed in an Omnifit glass column (25 mm × 400 mm, Diba Industries Ltd., Cambridge, UK) according to the manufacturer's protocol. The resulting bed volume was 13.4 mL. A gradient elution was performed at a flow rate of 0.7 mL/min with 25 mM sodium phosphate buffer pH 7.2 as mobile phase. After column equilibration, 50 mL sample mixture was loaded onto the column. Elution was performed with a gradient from 0 to 200 mM sodium chloride over 21 column volumes. The fractionation volume was set to 5 mL.

2.2.3. Molecular weight determination

Peak fractions resulting from the CEX separation were analyzed with respect to lysozyme PEGamer sizes using combined size exclusion chromatography (SEC) and light scattering (LS). This analysis was conducted using an ÄKTA Ettan system from GE Healthcare (Uppsala, Sweden) in combination with a Dawn Heleos 8+multi-angle light scattering detector and an Optilap rEX refractive index (RI) detector, both from Wyatt Technology (Santa Barbara, USA). For SEC a Superdex 200 GL10/300 (GE Healthcare, Uppsala, Sweden) with a mobile phase of 25 mM sodium phosphate, pH 7.0, containing 150 mM NaCl was used. The flow rate was set to 0.8 mL/min. Injection volumes between 50 and 100 μL were chosen. After UV absorbance monitoring at 280 nm, each sample was measured by LS and RI detection. Molecular weight (M_w) and hydrodynamic radius (R_h) calculation were calculated using the ASTRA software (v.5.3.4.18).

2.3. Chromatography system setup

An UltiMate3000 RSLC ×2 Dual system from Dionex (Sunnyvale, CA, USA) was used for UHPLC analysis. The system was composed of two HPG-3400RS pumps, a WPS-3000TFC-analytical autosampler and a DAD3000RS detector. The autosampler was equipped with a 5 μL sample loop. The volume of the injection needle was 15 μL and the syringe size was 250 μL. For control of the UHPLC equipment and for data evaluation the Chromeleon software (6.80 SR10) was used. The software was extended by an additional time base. This enables a virtual separation of the LC system in two parts which can then be controlled separately. Such a setup is necessary in order to facilitate separate data recording of each analyzed sample when performing SEC in interlaced mode. A thorough description of this setup and the performance of interlaced chromatography has been described by Farnen et al. [9] and Diederich et al. [8].

2.4. Size exclusion chromatography

SEC columns (Zenix SEC-300) were purchased from Sepax Technologies (Newark, DE, USA). The Zenix SEC-300 phase is a silica based material with a hydrophilic coating. The 3 μm sized particles have a nominal pore size of 300 Å. Columns of the dimensions 4.6 mm × 150 mm and 4.6 mm × 300 mm were used. The short column was operated with a flow rate of 0.6 mL/min and the long column with a flow rate of 0.4 mL/min. The columns were mounted with 0.2 μm Opti-Solv® EXP™ inlet filters (Optimize Technologies, Oregon City, OR, USA). For analysis, 5 μL sample was injected via full loop injections and a 250 mM potassium phosphate buffer at pH 6.8 with 200 mM potassium chloride was used as running buffer. To prevent fouling of the columns due to PEG binding, 10% (v/v) analytical grade ethanol was added to the buffer. Interlaced injection mode was performed in order to eliminate lag time between injection and elution of the first sample components.

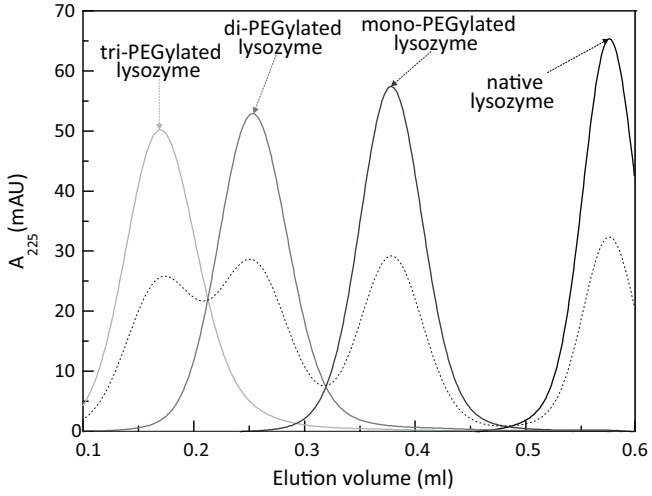

Fig. 1. Chromatograms used for calibration of the multivariate regression models. One chromatogram of each pure component with a concentration of 1 g/L with respect to lysozyme and one chromatogram of a mixture of native lysozyme and the three lysozyme PEGamers, all with a concentration of 0.5 g/L with respect to lysozyme.

2.5. Multivariate calibration

Where nothing else is stated, the multivariate regression models were calibrated with six chromatograms: single component chromatograms of each purified lysozyme PEGamer (mono-, di-, and triPEG) and native lysozyme, one chromatogram of all components in mass equivalent ratio with respect to lysozyme, and one chromatogram of a blank injection. Chromatograms of the pure components and the mixture of all four components are shown in Fig. 1A. The samples of pure components all had a concentration of 1 g/L. In the mixed sample each component had a concentration of 0.5 g/L. All mentioned concentrations are with respect to lysozyme.

Different multivariate regression types were applied: multilinear regression (MLR), partial least squares (PLS) regression, and multivariate curve resolution (MCR). All data processing for multivariate calibration was performed with MATLAB. For multivariate regression with MLR and PLS, the chromatographic data was preprocessed by mean centering. All PLS based regressions were based on 4 latent variables. MCR was performed with the MCR-ALS algorithm. After the MCR model was calibrated with the above-mentioned six chromatograms, it was able to deconvolute overlapping chromatograms unknown to the model. This resulted in values corresponding to the integrated area of each pure component present in each of the validation samples. From the initial MCR model calibration, the relation between area and concentration was known and based hereon, the concentration of each component in each validation sample was calculated. Hence, the MCR procedure is a combination of a multivariate approach for chromatogram deconvolution combined with an univariate calibration of the determined area and related concentration.

For multivariate calibrations based on integrated peak areas, calibration samples were generated according to a three layer onion design generated with MODDE (Umetrics, Sweden) which consisted of 32 combinations of the four components.

2.5.1. Chromatographic data

The multivariate calibrations were based either on full chromatograms or four defined points within the chromatogram. The four points were either defined as the UV signal at the elution volume corresponding to the peak maximum of each pure component or as the UV signal at actual peak maxima of the recorded chromatogram (see Fig. 2, left). For the purpose of comparison, calibrations based on integrated peak areas were also performed. Here, vertical peak limits as well as exponential rider skimming were applied to determine peak areas (see Fig. 2, right). Data from full chromatograms consisted of the UV the signal recorded with a frequency of 10 Hz for the short column and 5 Hz for the long column in a defined elution range. This range was defined as 1.4–1.9 mL for the short column and 2.8–3.75 mL for the long column with respect to chromatograms based on conventional injection mode as shown in Fig. 3. A full chromatogram consisted of 712 data points when generated with the long column and 500 data points when generated with the short column.

2.5.2. Determination of calibration precision

The calibrated concentration range for native lysozyme and each PEGamer was 0–1 g/L with respect to lysozyme. To determine the precision of this suggested approach for evaluation of chromatograms featuring low resolution, validation samples of defined concentrations were analyzed. The validation samples were prepared according to a three layer onion design (designed by MODDE) which resulted in 32 samples with seven different concentration levels (0, 0.25, 0.33, 0.5, 0.67, 0.75, and 1.0 g/L). From each sample three chromatographic runs were made and before each injection of a new sample, a blank run was performed. This procedure was performed using both the long and the short column. Based on the obtained chromatographic data, the calibrated multivariate models were used to predict the content of each lysozyme PEGamer and native lysozyme in the validation samples. For each model and each component the 95% confidence intervals were calculated based on the total of 96 analyzed samples (threefold injections of 32 validation samples). First the MATLAB 'poly1' fit function was used to fit a linear function to the relation between the concentration predicted by the multivariate model and the nominal concentration in the validation samples. After the fitting procedure, MATLAB was programmed to return the upper and lower confidence bounds for each linear fit. Finally, the confidence interval for each calibration and component was calculated by adding the upper and lower

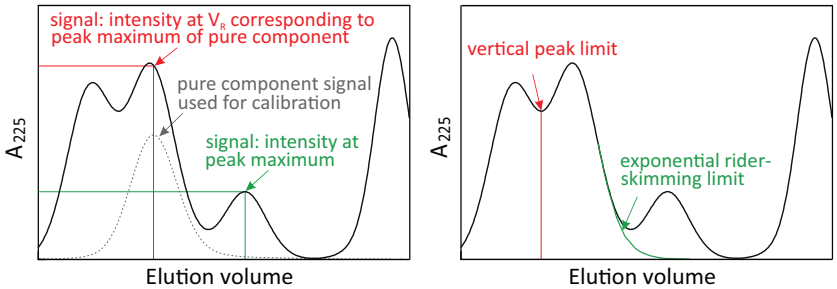

Fig. 2. Schematic illustration of the different chromatographic data used for calibration. Left: definition of peak maxima for calibration based on single chromatogram data points. Right: definition of peak limits for calibration based on area integration.

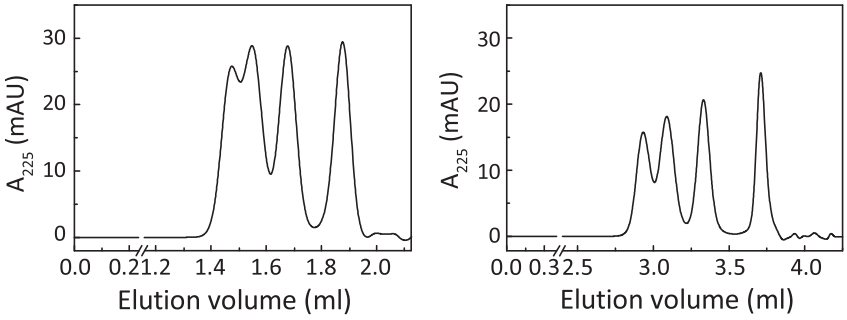

Fig. 3. Separation of native lysozyme, mono-, di- and, tri-PEGylated lysozyme with two columns of different lengths. The samples injected on the two columns were identical and contained equal amounts of each component with respect to lysozyme. Left: separation performed on a 150 mm column. Right: separation performed on a 300 mm column.

confidence bound. It should be noticed that the confidence interval does not to give exact information on the precision for different concentration levels within the calibration. Hence, the confidence intervals were only used for comparison of the different calibrations. To obtain more detailed information on precision, the relative standard deviation (RSD%) was calculated for each concentration interval and presented for a selection of the most precise calibrations.

2.5.3. Dynamic and static calibration

Two modes of multivariate calibration were used. The simplest was a static mode where one set of calibration sample chromatograms recorded immediately before analysis of the validation samples was used for calibration. However, to be able to account for systematic changes in column performance over time, a dynamic approach was developed. In order to do so, calibration sample chromatograms were recorded both before and after the analysis of the validation samples. The regression model was then recalibrated for each validation sample based on a linear interpolation between the chromatograms recorded before and after analysis of the validation samples. The interpolations were calculated using Eq. (1):

$$v_{i,j} = \frac{n-j}{n} \cdot v_{i,pre} + \frac{j}{n} \cdot v_{i,post} \quad (1)$$

where v_i is a vector containing the recorded chromatogram of calibration sample i, n is the total number of measured samples, and j is the number of the sample in the sample sequence for which the model is recalibrated. The indices *pre* and *post* indicates whether a calibration sample was measured before or after the validation samples.

3. Results and discussion

The aim was to establish a fast SEC assay (~2 min) for the quantification of lysozyme PEGamers and native lysozyme. Hence, a type of SEC columns which allows the use of rather high flow rates up to 3 cm/min was applied. In order to eliminate the lag time inherent in SEC analysis, injections were performed in interlaced mode. Finally, the narrow diameter of the chosen column type allowed for a sample size of 5 μL which minimizes the time necessary for sample injection preparation by the autosampler. All these measures resulted in an assay time of 5 and 2 min for the long and the short column, respectively. These measures, however, also caused insufficient resolution for precise quantification of all the lysozyme PEGamers when based on determined peak areas. Therefore, multivariate calibration was applied to achieve accurate and precise quantification despite the low resolution.

3.1. Assay time and chromatographic resolution

The resulting separation of lysozyme PEGamers using two different column lengths (150 and 300 mm) is shown in Fig. 3. The presented chromatograms result from samples injected in traditional sequential mode and, as can be seen, less than 50% of the recorded signal contains relevant information. Hence, interlaced injection mode was applied to reduce the analysis time significantly without decreasing the resolution further. The final assay for the long column was performed with an elution volume of 2.0 mL per injected sample at a flowrate of 0.4 mL/min. This resulted in an assay time of 5 min. The final assay for the short column was performed with an elution volume of 1.2 mL per injected sample at a flowrate of 0.6 mL/min. This resulted in an assay time of 2 min.

The determined chromatographic resolutions generated by both columns of lysozyme PEGamers and native lysozyme are listed in Table 1. These results are based on injection of equal amounts of the two components in question. A resolution of $R \gtrsim 1.5$ is in general sufficient for correct quantification based on integration of peak areas. The short column only generated sufficient resolution for native and mono-PEGylated lysozyme whereas the long column generated sufficient resolution for all components except the tri-PEGylated lysozyme. These results accentuate the challenging task of separating higher PEGylation forms using SEC to a degree sufficient for quantification based on area integration. If sufficient resolution is to be achieved for the higher PEGylation forms, a great amount of dispensable resolution will be generated for native and lower PEGylation forms. Therefore, in order to save time, multivariate calibration was applied in the attempt to achieve precise and accurate quantification despite low resolution.

3.2. Calibration precision

Multivariate models were calibrated based on multilinear regression (MLR), partial least squares (PLS) regression, and multivariate curve resolution (MCR). Further, traditional quantification based on univariate calibration of determined peak areas was performed in order to compare multivariate calibration with the traditional approach. Finally, MLR calibration based on peak areas and peak heights was applied to investigate whether low peak

Table 1
Chromatographic resolution listed for each column. Each resolution was determined by injecting equal amounts of the two components in question.

column length	chromatographic resolution			
	native	mono-PEG	di-PEG	tri-PEG
300 mm		2.72	1.47	0.79
150 mm		1.66	0.98	0.35

Table 2
Calibration results presented in order of precision. Different regression types were used to calibrate multivariate models based on different chromatogram data for evaluation of analytical chromatograms. The resulting confidence intervals are listed for each of the different regression types along with model mode and the applied chromatogram data. Further, results based on traditional area integration and univariate calibration are listed. See Fig. 2 for a schematic illustration of the different applied chromatogram data.

Regression type	Model mode	Data range	Confidence interval
300 mm column			
MLR	Static	Full chromatogram	0.0155
MLR	Static	Peak maxima[a]	0.0164
MCR	Static	Full chromatogram	0.0167
PLS	Static	Full chromatogram	0.0186
MLR	Static	Peak maxima[b]	0.0206
MLR	Static	Peak area[c]	0.0394
Univ.	Static	Peak area[c]	0.0476
Univ.	Static	Peak area[d]	0.0768
150 mm column			
MLR	Static	Peak maxima[a]	0.0226
MCR	Static	Full chromatogram	0.0226
MLR	Static	Full chromatogram	0.0247
PLS	Static	Full chromatogram	0.0347
MLR	Static	Peak area[c]	0.0453
MLR	Static	Peak maxima[b]	0.0491
Univ.	Static	Peak area[c]	0.0580
MLR	Dynamic	Peak maxima[a]	0.0164
MLR	dynamic	Full chromatogram	0.0167
MCR	Dynamic	Full chromatogram	0.0173
PLS	Dynamic	Full chromatogram	0.0191

[a] Signal height to base line at retention volumes corresponding to peak maxima of pure components.
[b] Signal height to baseline at peak maximum.
[c] Vertical peak limits.
[d] Exponential rider skimming.

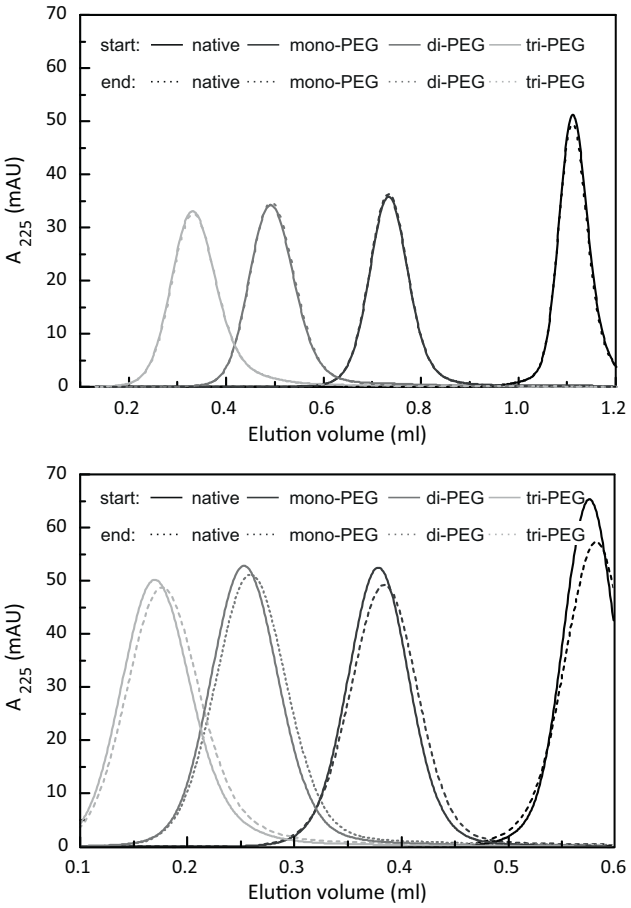

Fig. 4. Overlay of single component chromatograms before and after 140 sample injections. Top: 300 mm column. Bottom: 150 mm column.

resolution could be compensated by multivariate calibration of these parameters.

Confidence intervals were used to compare the precision of the different calibration approaches, both with regard to regression type and the data used for calibration (i.e. full chromatograms, peak maxima, peak areas). For this purpose, the mean confidence interval of the two components with the lowest chromatographic resolution was determined, in this case the di- and tri-PEGamers. The results are listed in Table 2 and are presented and discussed in the following sections.

3.2.1. Static calibration

In general, a clear increase in precision was achieved by applying static multivariate calibration of chromatographic data. For the long column, the precision increased from 0.0476 g/L to 0.0155 g/L by applying MLR to full chromatogram data instead of univariate calibration of peak areas. Based on equivalent calibration for the short column, an increase in precision from 0.0580 g/L to 0.0247 g/L was achieved. For the stronger overlapping peaks generated by the short column, MLR calibration based on only four points in the chromatogram (obtained at V_R of the pure components) resulted in a more precise calibration with a confidence interval of 0.0226 g/L. MLR gave more precise results than PLS and MCR both for the long and the short column. Further, MLR calibration based on peak areas was more precise than univariate calibration of peak areas, however much less precise than calibration based on full chromatogram data. MLR calibration based on peak heights at peak maxima was more precise than MLR calibrations based on peak areas for the long column, however the opposite was the result for the short column. The elution profiles of the pure components (see Fig. 4) shows that severe peak overlapping was limited to approximately 50% for the long column. Therefore the position of peak maxima was less influenced by the presence of other components when comparing the long column with the short column.

3.2.2. Dynamic calibration

The precision of the univariate peak area based calibration decreased by ~25% when comparing the results obtained by the long column to those obtained by the short column. In comparison, the precision of the multivariate calibration based on MLR and full chromatograms decreased by ~70%. Even though the absolute value of the determined confidence interval was still several times lower for the multivariate calibration compared to univariate calibration (0.0247 *vs.* 0.0580 g/L) when using the short column, the relatively high decrease in precision of the multivariate calibration when switching from the longer to the shorter column (*e.g.* 0.0155 *vs.* 0.0247 g/L) was not expected. In Fig. 4A an overlay of pure component chromatograms recorded before and after 140 sample injections on the long column is shown. The data shows stable column performance without a systematic change in peak height or retention volume over time. Fig. 4B displays the equivalent data for the short column and here a shift in retention time for all components along with peak broadening can be observed. To overcome the inaccuracy caused by the change in column performance, a dynamic calibration approach was used. This calibration was based on linear interpolation of calibration chromatograms recorded immediately before and after the analysis of the validation samples. It was chosen to apply linear interpolation as the observed change in retention time was linear. The use of dynamic calibration for the short column resulted in an assay of similar precision compared to the static calibration of the long column (0.0167 *vs.*

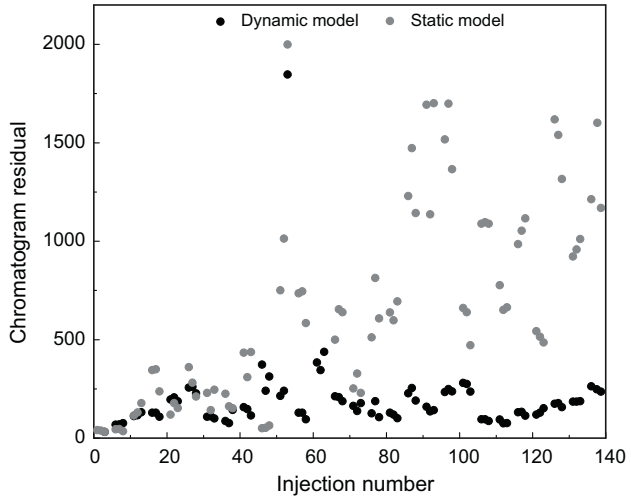

Fig. 5. Residuals of the 96 validation sample chromatograms as a function of injection order. Each gray data point represents one sample chromatogram evaluated the static PLS calibration. Each black data point represents one sample chromatogram evaluated with the dynamic PLS calibration. The residuals for the 32 blank samples recorded between the validation samples are not displayed.

0.0155 g/L – MLR, full chromatogram). In Fig. 5 the chromatogram residuals of the validation sample chromatograms are displayed. The shown residuals are derived from both the dynamic and the static PLS calibration for the short column based on the full chromatograms. The variation of the chromatogram residuals remains constant over the period of analysis for the dynamic calibration, where as the variation increases throughout the period of analysis for the static calibration. This supports the decision to use linear interpolation for the dynamic calibration.

In order to visualize the effect of dynamic calibration, three artificial chromatograms were created and displayed along with a true chromatogram of a validation sample (Fig. 6). The artificial chromatograms were created by a linear combination of pure sample chromatograms in a ratio corresponding to the true chromatogram.

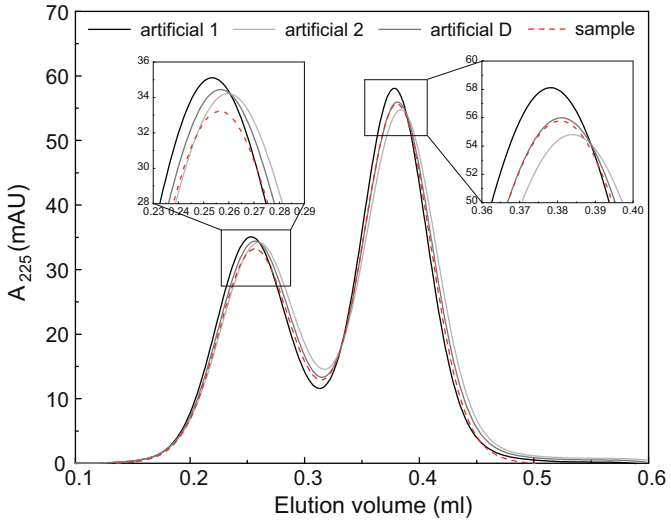

Fig. 6. Visual comparison of the effect of dynamic calibration. One true chromatogram and three artificial chromatogram based on linear combinations of pure component chromatograms are displayed. Two of the artificial chromatograms were built from the pure component chromatograms recorded either before (*artificial 1*) or after (*artificial 2*) analysis of the validation samples. The third artificial chromatogram (*artificial D*) was built from the linear interpolations used in the dynamic calibrations. The linear interpolation was built to correspond to the point in time where the true samples was analyzed.

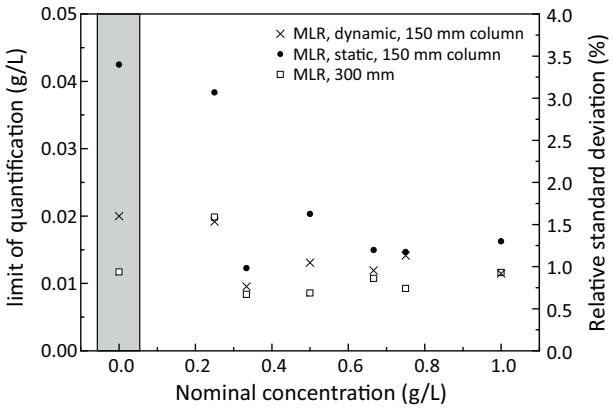

Fig. 7. RSD% values for the concentration levels included in the validation sample design. Displayed are mean values of the determined RDS% for di- and tri-PEGylated lysozyme. The values displayed for the concentration zero marked by the gray bar are LOQ values. Also the LOQ values are mean values of the determined LOQ for di- and tri-PEGylated lysozyme.

Two of the artificial chromatograms were based on the calibration chromatograms recorded either before or after the analysis of the validation samples. Neither of these chromatograms were of good resemblance to the validation sample chromatogram. The third artificial chromatogram was based on the linear interpolated chromatograms of the calibration samples which were created for the dynamic calibrations. This chromatogram was built to match the point in time where the displayed true chromatogram was recorded and exhibited a high resemblance to the true validation sample chromatogram. For the right peak (monoPEG–protein conjugate), the resemblance was high both for peak intensity and peak shift. The fact that the left peak of the sample chromatogram deviated from the dynamic calibration chromatogram in peak intensity is most likely due to a pipetting error during preparation of the validation samples.

3.2.3. Concentration related precision and sensitivity

The concentration related precision was determined for the dynamic and static MLR calibrations based on the short column and the MLR calibration based on the long column. The relative standard deviation (RSD%) for each concentration is displayed in Fig. 7. The displayed RSD% values are mean values of the RSD% for di- and tri-PEGylated lysozyme. For all three calibrations, the precision increased from the first to the second concentration level (0.25–0.33 g/L). For all concentration levels, the precision of the static MLR calibration based on the 150 mm column was lowest (RSD ~ 3–1.5%). The precision for all concentration levels was similar for the dynamic MLR calibration based on the 150 mm column and the static MLR calibration based on the 300 mm column. This is in accordance with the finding that the overall precision of these calibrations were similar (95% confidence interval: 0.0167 g/L and 0.0155 g/L – see Table 2). This means that both the overall precision and the specific precision achieved for the long column can be maintained for the short column by applying multivariate calibration. For further comparison and characterization of the calibrations, the limit of quantification (LOQ) was determined. LOQ was determined as six times the standard deviation of all 'zero' predictions in the calibration samples for each component. Again, the displayed LOQ values are mean values of the determined LOQ value for di- and tri-PEGylated lysozyme. With regard to the LOQ value, the dynamic approach does increase the sensitivity of the method, however not to the level of the MLR calibration based on the 300 mm column.

4. Conclusion and outlook

The aim of establishing an SEC assay for the quantification of lysozyme PEGamers with an analysis time of 2 min was achieved. The presented work clearly demonstrates that by applying multivariate calibration for the quantitative evaluation of low resolution chromatograms the precision can be enhanced significantly when compared to traditional univariate calibration. Hence, fast chromatographic assays can easily be achieved by applying short columns and fast flow rates. The tested chromatographic assays included resolutions down to 0.35 demonstrating how little resolution is actually sufficient to achieve a highly precise chromatographic assay. The lower limit of resolution necessary for precise calibration is still to be determined. Further, the results demonstrated that a change in column performance over time can be handled without difficulty by using a dynamic calibration approach.

In the presented work, pure samples were used for calibration. If pure material is at hand or easily achievable, this is the most straightforward approach. If pure samples are difficult or impossible to obtain, mixed samples of defined composition can also be used. This then requires an alternative assay or analytics to define the mixed samples and further these samples must contain sufficient variation.

The use of multivariate calibration is of course not limited to SEC. Any robust chromatographic assay established for defined samples can be evaluated by the presented approach. The approach cannot be applied to complex samples in which unknown peaks occur in the area used for calibration and also chromatographic assays based on non-linear chromatography will demand more sophisticated multivariate calibration of non-linear nature. This might increase the calibration complexity to a level which is no longer leveraged by the gained increase in assay speed. However, this is still to be investigated.

In working environments with limited time for assay development and data evaluation, it might be an unmanageable task to perform data evaluation not inherent in the chromatography system software. Hence, we propose the integration of multivariate calibration directly in the commercial software supplied with the chromatography systems.

References

[1] M.J. Knauf, D.P. Bell, P. Hirtzer, Z.P. Luo, J.D. Young, N.V. Katre, J. Biol. Chem. 263 (29) (1988) 15064.
[2] F. Fuertges, A. Abuchowski, J. Control. Release 11 (1990) 139.
[3] N.V. Katre, M.J. Knauf, W.J. Laird, Proc. Natl. Acad. Sci. USA 84 (6) (1987) 1487.
[4] R. Bhambure, K. Kumar, A.S. Rathore, Trends Biotechnol. 29 (3) (2011) 127.
[5] S. Oelmeier, F. Dismer, J. Hubbuch, Biotechnol. Bioeng. 108 (1) (2011) 69.
[6] S. Chhatre, R. Francis, D.G. Bracewell, N.J. Titchener-Hooker, J. Chromatogr. B 878 (30) (2010) 3067.
[7] M. Bensch, P.S. Wierling, E. von Lieres, J. Hubbuch, Chem. Eng. Technol. 28 (11) (2005) 1274.
[8] P. Diederich, S.K. Hansen, S.A. Oelmeier, B. Stolzenberger, J. Hubbuch, J. Chromatogr. A 1218 (50) (2011) 9010.
[9] D. Farnan, G.T. Moreno, J. Stults, A. Becker, G. Tremintin, M. van Gils, J. Chromatogr. A 1216 (51) (2009) 8904.
[10] C.J. Fee, J.M. Van Alstine, Chem. Eng. Sci. 61 (3) (2006) 924.