



Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles

Julian Kopp^{a,*}, Florian B. Zauner^b, Andreas Pell^a, Johanna Hausjell^c, Diana Humer^c, Julian Ebner^c, Christoph Herwig^a, Oliver Spadiut^c, Christoph Slouka^c, Reinhard Pell^b

^a Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, TU Wien, Gumpendorfer Straße 1a, 1060 Vienna, Austria

^b SANDOZ GmbH, Mondseestrasse 11, 4866 Unterach, Austria

^c Research Division Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Gumpendorfer Straße 1a, 1060 Vienna, Vienna, Austria

ARTICLE INFO

Article history:

Received 27 February 2020

Received in revised form 13 May 2020

Accepted 2 June 2020

Available online 4 June 2020

Keywords:

Liquid chromatography

(U)HPLC

Protein quantification

E. coli

Analytical quality by design (AQbD)

Design of Experiment (DoE)

ABSTRACT

Biopharmaceutical drug substances are generally produced using fermentation technology and are subsequently purified in the following downstream process. For the determination of critical quality attributes (CQAs), such as target protein titer and purity, monitoring tools are required before quality control analysis. We herein present a novel reversed phase liquid chromatography method (RPLC), which enables facile and robust protein quantification during upstream and downstream processing of intracellularly produced proteins in *E. coli*.

The overall goal was to develop a fast, robust and mass spectrometry compatible method which can baseline resolve and quantify each protein of interest. Method development consisted of three steps, oriented on an Analytical Quality by Design (AQbD) workflow: (i) the stationary phase as primary parameter was chosen based on state-of-the-art technology thus minimizing protein on-column adsorption and providing high efficiency, (ii) secondary parameters (i.e. gradient conditions and column temperature) were optimized applying chromatographic modeling, and (iii) the established Method Operable Design Region (MODR) was challenged and confirmed during robustness testing, performed *in-silico* and experimentally by a Design of experiment (DoE) based approach.

Finally, we validated the RPLC method for pivotal validation parameters (i.e. linearity, limit of quantification, and repeatability) and compared it for protein quantification against a well-established analytical methodology. The outcome of this study shows (i) a protocol for RPLC development using an AQbD principle for new method generation and (ii) a highly versatile RPLC method, suited for quick and straightforward recombinant protein titer measurement being applicable for the detection of a broad range of proteins.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations: CQAs, critical quality attributes; RPLC, reversed phase liquid chromatography; AQbD, Analytical Quality by Design; MODR, Method Operable Design Region; DoE, Design of experiment; KPI, key performance indicator; *E. coli*, *Escherichia coli*; IB, inclusion body; SDS-PAGE, sodium-dodecyl-sulfate polyacrylamide gel electrophoresis; LC, liquid chromatography; CE, Capillary electrophoresis; MS, mass spectrometry; HIC, Hydrophobic interaction chromatography; SEC, size exclusion chromatography; IEX, ion-exchange chromatography; mAbs, monoclonal antibodies; ADCs, antibody-drug conjugates; TFA, trifluoroacetic acid; SPP, superficially porous particles; N-pro, N-pro fusion protein; GFP, green fluorescent protein; CH3H, Chalcone 3-Hydroxylase; HRP, Horseradish peroxidase; BSA, Bovine Serum Albumin; MW, Molecular weight; GRAVY, grand average of hydropathicity; FLD, fluorescence detector; UV, ultraviolet; EP, European Pharmacopeia; tG, gradient time; ATP, analytical target profile; R², measure of fit; R², adj. adjusted measure of fit; Q², predictability; LOQ, limit of quantification; RSD, relative standard deviation.

* Corresponding author.

E-mail address: julian.kopp@tuwien.ac.at (J. Kopp).

<https://doi.org/10.1016/j.jpba.2020.113412>

0731-7085/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A panoply of commercially available biopharmaceuticals are manufactured as recombinant proteins using microbial or mammalian hosts [1]. To monitor the production of biopharmaceutically active compounds, product titer is a pivotal key performance indicator (KPI) [2]. Since it is necessary to react promptly on process changes, facile sample preparation and subsequent rapid analytics are of high interest [3]. The gram-negative bacterium *Escherichia coli* (*E. coli*) still is today's expression host of choice for production of approximately 30 % of recombinantly produced pharmaceuticals in industry [1]. Quantification of target proteins is challenging in *E. coli* as recombinant proteins generally are expressed in the cell interior (either soluble or as an inclusion body = IB). Necessary cell disruption leads to a high amount of host-cell associated impurities and thus to a complex sample matrix for analysis.

Today, a plethora of analytical technologies exist for the analysis of (recombinant) proteins. Favored techniques are sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which can be further used in western blotting to detect low protein concentrations, or capillary electrophoresis [4]. Samples derived from microbial hosts often suffer from impurities, favoring target protein detection via size using SDS-PAGE analysis [5]. Different stains, such as silver and comassie blue, can be employed for protein staining of gels. Silver staining PAGE enables more sensitive staining, however comassie blue SDS-PAGE is used more frequently as it is easier in its application, cheaper and also allows much faster processing compared to silver staining [5]. Still, when comparing SDS-PAGE analysis to liquid chromatography (LC) it is more time demanding and lacks accurate quantitation [6]. Capillary electrophoresis (CE) is another tool for the qualitative and quantitative analysis of biopharmaceuticals, and proteins in general [7]. CE is considered a powerful separation technique in terms of selectivity and efficiency, and can further be coupled with mass spectrometry (MS) for protein identification and characterization [8,9]. However, it has drawbacks such as limited instrument ruggedness and non-straightforward method development.

In the last decade, LC has emerged as technique of choice for characterization and quantitation of (recombinant) proteins [10–13], due to facile sample preparation, technically mature instrumentation and major advancements in stationary phase chemistry. Hydrophobic interaction chromatography (HIC) [14], size exclusion chromatography (SEC) [15] and ion-exchange chromatography (IEX) [11] enable analysis of proteins in native (non-denaturing) conditions, but are limited in their applicability. SEC is mainly carried out for quantitation of protein aggregates and IEX for separation of protein charge variants. HIC (on analytical scale) has its main application in separation and characterization of large biomolecules, such as monoclonal antibodies (mAbs) and antibody-drug conjugates (ADCs) [16]. However, HIC exhibits drawbacks like slow mass transfer resulting in broad peaks and the necessity of high mobile phase salt concentrations thus impeding hyphenation with MS.

Reversed phase liquid chromatography (RPLC) has become the predominant technique for analysis of proteins due to its versatility, robustness, high efficiency and MS compatibility [17,18]. In contrast to the aforementioned chromatographic modes, RPLC is a denaturing technique which often requires harsh conditions (e.g. high temperature and low pH) in order to eliminate irreversible on-column protein adsorption and improve peak shape. However, the advent of novel stationary phases, including wide-pore superficially porous particles (SPP) and special stationary phase coatings for suppression of silanol interactions, enabled higher peak efficiencies and application of milder elution conditions [19,20]. Recently, a wide-pore SPP column with a high coverage phenyl bonding

was introduced to the market, which addresses the last obstacles in RPLC protein analysis (i.e. on-column adsorption of large biomolecules, poor selectivity for closely related proteins) [21].

In this study, we established a widely-applicable, robust, and MS-compatible RPLC method applicable for on-line process analytics and quality control of recombinant proteins with pronounced differences in their physicochemical properties. A state-of-the-art column was used for systematic method development based on Analytical Quality by Design (AQbD) principles. The needed requirements to the method were defined according to the ATP (= analytical target profile), followed by investigation of critical method parameters and their effects onto monitored responses using a Design-of Experiment (=DoE) approach [22]. Following the AQbD concept, the method operable design region (MODR) including retention modeling was assessed [23]. Robustness testing was performed *in-silico* and experimentally to verify the set MODR, applying a Design-of-Experiment (DoE) concept. Finally, the method performance was compared for protein quantification in a case study against one of the benchmark technologies (i.e. SDS-PAGE).

2. Material and methods

2.1. Chemicals and samples

Ultra-pure water was obtained from a Milli-Q system from Merck Millipore (Burlington, USA). Acetonitrile (HPLC grade) and trifluoroacetic acid (TFA, >99.9 %) were purchased from Carl Roth (Karlsruhe, Germany). All proteins implemented within this study are described below. Further information regarding cultivation and purification of the proteins can be found in Supplementary section A.

2.1.1. N-pro fusion protein (N-pro)

N-pro fusion protein was produced in an *E. coli* BL21(DE3) with the pET30a plasmid system (kanamycin resistance). The target protein was linked to a N-pro fusion protein used for purification. Details on protein structure and purification steps cannot be given due to proprietary reasons.

2.1.2. eGFP (GFP)

GFP was produced in an *E. coli* BL21(DE3) (Life technologies, Carlsbad, CA, USA), transformed with a pET21a(+) vector (ampicillin resistance) carrying the gene coding for enhanced green fluorescent protein (eGFP). Samples derived from fed-batch fermentation (as referred here [24]), were disrupted and centrifuged to receive the soluble fraction. Protein solution was filtered through pores < 0.2 µm and was further purified with an immobilized metal affinity chromatography (IMAC), binding to the expressed HIS-tag of the protein. Samples were desalted and stored in aqueous buffer composed of 20 mM TRIS, 100 mM NaCl, pH = 7.5 containing 10 % glycerol for RPLC analysis.

2.1.3. Chalcone 3-Hydroxylase (CH3H)

CH3H was produced in an *E. coli* HSM174(DE3) genotype: F- recA1 hsdR(rK12- mK12+) (DE3) (Rif^r) in a pET21a plasmid (Novagen, Madison, WI, USA) encoding CH3H of *D. variabilis*. The plasmid for CH3H expression was a kind donation from Christina Divne and Rosaria Gandini (Karolinska Institute, Stockholm, Sweden). In brief: sub-cloning of the CH3H sequence into pNIC-CTHO was performed using the LIC-cloning methodology. Inclusion body samples (derived from fed-batch fermentation) were analyzed as crude samples after homogenization and samples were solubilized with a buffer described elsewhere [24].

Table 1
Physicochemical properties of analyzed proteins.

Target protein	MW [kDa]	Disulphide bridges [nr]	GRAVY Index [25]	expression form
N-pro	28.0	n.a.	n.a.	IB
BSA	66.5	17	−0.433	n/a
GFP	26.9	1	−0.467	soluble and IB
HRP	34.7	4	−0.177	IB
CH3H	55.0	0	−0.080	IB

2.1.4. Horseradish peroxidase (HRP)

HRP (EC 1.11.1.7) was produced in an *E. coli* BL21(DE3) pET21d(+) with ampicillin resistance (AmpR). The *hrp*-gene coding for HRP variant C1A was codon-optimized for *E. coli* and obtained from GenScript USA Inc. (Piscataway, NJ, USA). Samples derived from fed-batch fermentation were disrupted and centrifuged. The aggregated fraction was further purified [22]. Inclusion bodies were washed with water, solubilized, refolded and purified using a yet unpublished procedure. Samples for analysis were stored in an aqueous buffer containing 20 mM BIS-TRIS and 7% glycerol.

2.1.5. Bovine serum albumin (BSA)

BSA was purchased from Sigma-Aldrich (St. Louis, MO, USA), receiving 2 mg/mL aliquots for analysis stored in 0.9 % NaCl containing 0.05 % sodium azide (CAS Number 9048–46–8; MDL number MFCD00130384).

2.1.6. Protein physicochemical properties

Physical properties and number of disulfide bonds were calculated using different software tools and are summarized in Table 1. Molecular weight (MW) and grand average of hydropathicity (GRAVY) were calculated using ExPASy ProtPar [25] as these features can be accessed on a free basis with the given amino acid composition of the target protein. Disulfide prediction was calculated using the DiANNA 1.1 webserver by F. Ferre et C. Clote [26].

2.2. Equipment and software

Measurements were performed using a Dionex UltiMate 3000 system with a quaternary solvent delivery pump, an auto-sampler, UV and fluorescence (FL) detector (Thermo Fisher, Waltham, MA, USA). Dwell volume was determined as $V_d = 0.90$ mL. Data were acquired using either an UV detector or a FL Detector. Data acquisition and instrument control were carried out by Chromeleon 7.2 software (Thermo Fisher). Calculations and data transfer were achieved with Excel (Microsoft). Resolution was calculated according to European Pharmacopeia (EP). Retention and resolution modeling was performed with DryLab® software (Version 4.3.3., Molnár-Institute, Berlin, Germany). DoE and statistical evaluations were carried out with the software MODDE (MODDE 12.1, Umetrics, Sweden).

2.3. Chromatographic conditions

The HPLC was equipped with a BioResolve RP mAb Polyphenyl column (dimensions 100×3 mm, particle size $2.7 \mu\text{m}$) which is designed for mAb and large protein analyses (Waters Corporation, MA, USA). In order to prolong column lifetime, a pre-column (3.9×5 mm, $2.7 \mu\text{m}$) was used. The mobile phase was composed of ultrapure water (=MQ, eluent A) and acetonitrile (eluent B) both supplemented with 0.10 % (v/v) trifluoroacetic acid. The injection volume was set to $2.0 \mu\text{L}$. The flow rate and the column temperature of the final method were 0.4 mL/min and 70°C , respectively. Re-equilibration was conducted for six minutes, which corresponds to approximate five column volumes. The acquisition wavelength for UV detection was set to 214 and 280 nm, respec-

tively, for the FL detection the excitation wavelength was set to 280 nm and the emission wavelength to 360 nm. All detectors were series-connected, monitoring UV and FLD chromatograms at once, although FLD chromatograms only were used for evaluation. Sample concentration was 1.5 mg/mL for BSA. Via BSA calibration the sample nominal concentration was determined to be 2.5 mg/mL for N-pro, 3.5 mg/mL for GFP, 0.5 mg/mL for HRP and 2.0 mg/mL for CH3H. Total observed values in FLD- spectra differ as the amount of aromatic amino acids highly varies between employed target proteins.

2.4. Design of experiment of input runs for software modeling

For the chromatographic modeling using DryLab, a 2D model was chosen in order to optimize gradient time (tG) and column temperature (T). The input runs required a 2×2 full factorial design applying a linear gradient with short (tG of 8 min) and long (tG of 24 min) gradient time (starting from 25 % B to 60 % B), and column temperature of 50°C and 78°C , respectively (the concept is shown in Fig. 1a). The flow rate for the input runs was set to 0.8 mL/min . All runs were carried out separately for each protein and performed in duplicate. The mean value of the retention time was used for modeling in DryLab (the raw data are given in the Supplementary data section B, Table B1).

2.5. Protein assay determination (quantification)

For SDS-PAGE analysis BSA was mixed with 2-fold concentrated Laemmli solution to achieve single concentrated Laemmli buffer in the final dilution. Samples were then heated at 95°C for 10 min. $10 \mu\text{L}$ of each sample were loaded onto pre-cast SDS gels containing 15 vials (8–16 %, Bio-Rad, Hercules, CA, USA). Gels were run in a Mini-PROTEAN Tetra System (Bio-Rad, Hercules, CA, USA) for about 60 min at 140 V and stained with Coomassie Blue. The protein bands were evaluated densitometrically using the software ImageLab (Bio-Rad, Hercules, CA, USA).

For HPLC-measurement of the soluble proteins, the sample was directly used after cell disruption and centrifugation for GFP and after a refolding step for HRP [24]. IB pellet samples were prepared according to Ref. [24], and subsequently solubilized using a buffer solution of 7.5 M guanidine hydrochloride, 62 mM TRIS, and 125 mM 1,4-Dithioerythritol (DTT) at pH = 8 (all chemicals purchased from Carl Roth, Karlsruhe, Germany). All samples were filtered through $0.2 \mu\text{m}$ syringe filters purchased from Carl Roth (Karlsruhe, Germany) prior to analysis.

2.6. Validation

As BSA was stored in an aqueous solution, dilutions of the commercial purchased standard were done with ultra-pure water (Burlington, USA). Solutions for LOQ and linearity determination consisted of the following concentrations: (5, 10, 50, 100, 250, 500, 1000, 1500, and 2000 $\mu\text{g/mL}$). LOQ was determined experimentally by a six-fold injection followed by evaluation of the relative standard deviation of peak areas and the S/N ratio. Linearity was evaluated in the range of determined LOQ until 2.0 mg/mL (=con-

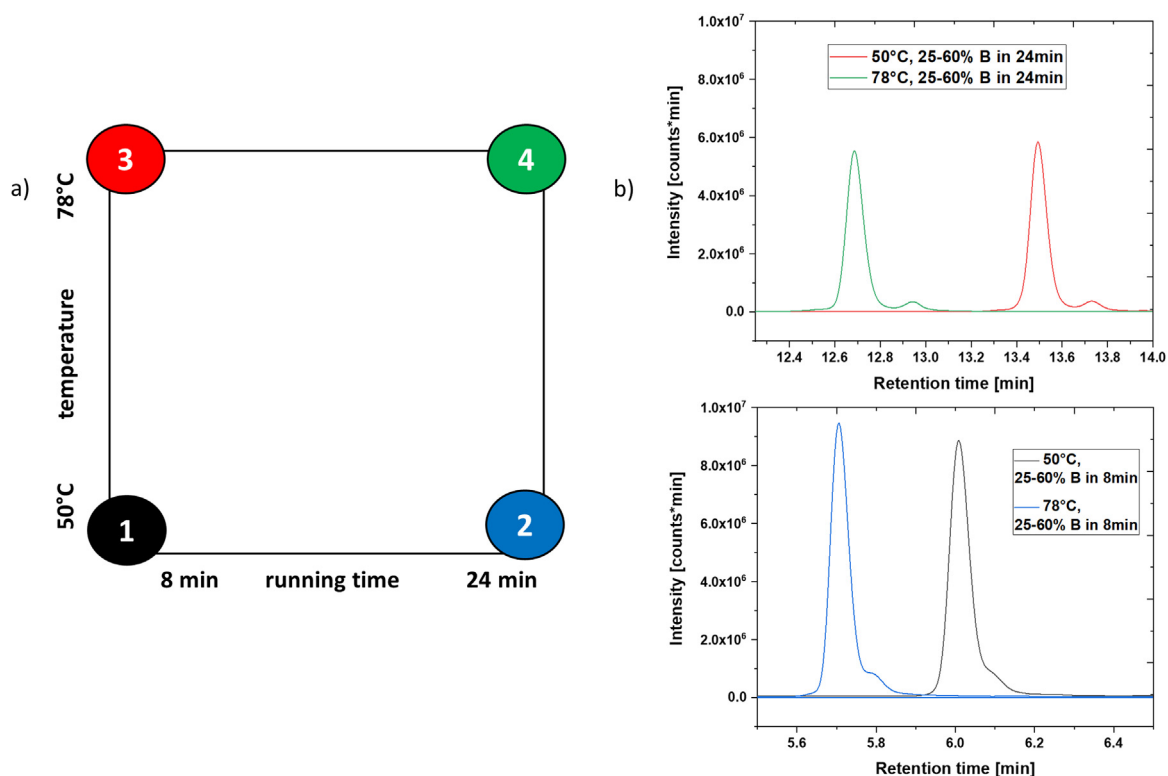


Fig. 1. a) DoE for secondary method factors varied regarding their input conditions for modeling software DryLab4 operated at a constant flow rate of 0.8 mL/min; b) Exemplary Input chromatograms for the individual DOE points, illustrated by the model protein HRP.

centration of the commercially available reference standard) at nine concentration levels.

3. Results and discussion

In order to develop a widely applicable LC method for protein assay determination, five model proteins were chosen which cover a broad range of protein size and hydrophobicity, as hydrophobicity is known to influence RPLC majorly. The GRAVY Index, describing hydrophobicity of target proteins, is calculated via amino acid composition of target proteins (listed in Table 1): proteins resulting in a GRAVY index close to 0 are highly hydrophobic whereas proteins are more hydrophilic, the more negative results for GRAVY Index are [25]. In biotechnology, GFP is often utilized as biosensor or as a marker protein due to easy auto-fluorescent detection [2]. BSA is often used for calibration purposes in biotechnology if there is a lack of a purified reference substance [24]. HRP, being used for enzymatic cancer treatment, was recombinantly expressed as an inclusion body but analyzed in its native form after refolding within this study. To show the applicability of the developed method independent from expression form, the two further analytes (N-pro fusion protein and CH3H) were analyzed in their incorrect folded form (=IB). All samples were evaluated separately as they were (i) derived from different purification steps (stated in Section 2.1. for each given protein) (ii) and were stored in different buffer solutions (stated in Section 2.5).

3.1. Method development workflow and selection of primary factors

The analytical target profile (=ATP) was established to develop a robust and sensitive method with baseline-separation between neighboring target peaks (i.e. critical resolution greater 1.5). As samples derived from microbial cultivation might cause a highly

complex sample matrix, the focus of this study was to provide a generic RPLC method for protein quantification, being independent from state of purification and its needed buffer solution. To verify this approach, target proteins were chosen in a broad range of hydrophobicity and varied in their degree of purity, as crude samples (i.e. CH3H) and highly purified proteins (i.e. HRP) were evaluated. MS-compatibility was further chosen as target profile in order to make the developed method applicable for further elucidation possibilities. The first step was the selection of appropriate method factors (parameters) for the stationary- and mobile phase system (stationary phase and mobile phase bulk solvents are considered primary method factors in an "AQbD language" [21,27]): Considering the recent advancements in stationary phase chemistry for RPLC protein analytics, a small set of columns is commercially available which minimize protein on column adsorption due to either special end-capping of the silica surface or using an organic polymer as support material [17,21]. Further, high peak efficiency can be accomplished on account of smaller particle sizes and/or usage of superficially porous particles [19,21]. Due to aforementioned reasons, a BioResolve RP mAb Polyphenyl column was chosen.

Apart from the stationary phase chemistry, also the type of organic modifier and mobile phase pH are considered as primary factors [20,27]. As primary factors are known to exhibit a higher impact on selectivity than secondary factors (i.e. column temperature), appropriate choice is highly important [21,27]. In contrast to RPLC analyses of small organic molecules, Acetonitrile (being an aprotic solvent) is most often applied as bulk mobile phase in protein RPLC [17]. The application of protic solvents (e.g. Methanol) often creates protein adsorption, hence they are avoided or only added in small portions to Acetonitrile in order to improve selectivity. For the sake of simplicity and in order to develop an easy to use method, Acetonitrile was chosen as the only organic modifier. Hence, utilization of purified water and acetonitrile were selected

Table 2

DryLab optimized method parameters.

Analysis time [min]	Eluent B [%]
0.0	25
8.0	55
10.0	75
12.0	75
12.1	25
18.0	25

Flow rate: 0.4 mL/min.**Column temperature:** 70 °C.

as mobile phase bulk solvents and TFA was used as additive to suppress secondary interactions. Also, TFA as “semi-volatile” additive enables hyphenation with MS or aerosol-forming universal detectors (i.e. charged aerosol or evaporative light scattering detectors), although detector sensitivity is compromised.

Gradient (tG) and column temperature (T), have been described to significantly enhance protein separation in RPLC [17] and thus were chosen as secondary factors. High temperature set-points were chosen to (i) decrease the chance of non-specific protein adsorption and (ii) to favor thermodynamic behavior for chromatographic interaction. Since TFA was applied because of aforementioned reasons, a systematic variation and optimization of pH was omitted.

3.2. DryLab assisted modeling of secondary method factors

Gradient time and column temperature were optimized using a chromatographic modeling software (i.e. DryLab [28,29]). The calculation of “ideal” parameters for the flow rate, gradient and column temperature required retention time data obtained from input runs. In this study, a 2×2 full factorial design was applied (illustrated in Fig. 1a). tG was varied by three-fold and T was set to 50 °C for the lower limit and to 78 °C for the upper limit (Fig. 1). Limits were chosen as temperatures below 50 °C may cause irreversible adsorption of proteins on the column, whereas 80 °C was the technical limit of the employed HPLC system (thus 78 °C was chosen to allow flexibility during robustness experiments). The changes in retention time are exemplarily displayed for HRP in Fig. 1b. Supplementary Table B1 lists each retention time per protein and chromatographic condition.

Based on the obtained retention times and peak-widths (at 50 % peak height) for each condition and protein, a chromatographic model was created in DryLab. The resolution between GFP and BSA (compare Fig. 2b and c) was identified as the most critical peak pair due to their congeneric hydrophobicity [25]. By iteratively modifying the method parameters *in-silico*, a constant flow rate of 0.4 mL/min, a column temperature of 70 °C and the gradient program shown in Table 2 were determined thus establishing the method operable design region (MODR).

Fig. 2a depicts the predicted chromatogram. In order to proof the validity of the DryLab model and its prediction, a verification run was carried out in the lab (Fig. 2b). As can be deduced by comparing the predicted with the real chromatogram, the retention times and peak widths are well comparable. Thus, the DryLab model is considered to be valid and can be used for further studies.

The critical resolution, defined as the smallest calculated resolution for all peak pairs, is displayed as a heat map in dependence of the column temperature versus the gradient time (Fig. 2c). This additionally helps in the selection of the optimal combination of column temperature and gradient time by simultaneously obtaining an estimate for the robustness of the selected method factors. In Fig. 2c, the red to orange area is the result of a combination of T and tG which leads to a critical resolution bigger than 2.0. The cross shows the selection of tG and T of the optimized method. Tempera-

Table 3List of all factors altered including their set-points and their variance, as well as the variance deviation for the *in-silico* calculation.

Factor	Set-point + variance delta
Flow rate [mL/min]:	0.4 ± 0.1
Dwell Volume [mL]:	0.9 ± 0.7
Temp. [°C]:	70 ± 5
Start %B [%]:	25 ± 1
Step 1 Time [min]:	8 ± 1
Step 1 %B [%]:	55 ± 1
Step 2 Time [min]:	10 ± 1
tG [min]:	12 ± 1
End %B [%]:	75 ± 1

tures below 70 °C could facilitate unspecific protein adsorption and were therefore neglected. The calculated chromatographic space around our set point is predicted to be in major parts characterized by a critical resolution > 2.0 and thus considered as robust by simultaneously minimizing the run time. Thus, the predicted method operable design region is equal to the area colored in red-orange in Fig. 2c.

3.3. In-silico and experimental robustness studies

3.3.1. In silico robustness

Robustness evaluates the capacity of an analytical method of remaining unaffected by small, but deliberate changes of method factors (parameters) [30]. DryLab offers the possibility of *in-silico* robustness studies [28]. Nine factors in total (compare to Table 3) were varied at three levels (central point, upper and lower level) and the critical resolution was calculated for each factor combination.

A total of 13122 operating conditions were simulated and tested for their influence onto the critical resolution. Remarkably, the outcome of the *in-silico* calculation is showing that a 94.1 % of all factor combinations were above the set resolution of 2.0, indicating a highly robust method. Moreover, the *in-silico* robustness study allowed to identify pivotal method factors with the highest impact on the critical resolution (compare to Supplementary data B, Figure B1). The factor exhibiting the highest influence identified by the *in-silico* robustness study was found to be the gradient time (compare to Supplementary Fig. 1B). Further factors, exhibiting high impact on the critical resolution were identified throughout simulation and hence chosen for experimental robustness determination as described in chapter 3.3.2.

3.3.2. Experimental determination of robustness

Due to the fact that the critical peak pair (i.e. BSA and GFP) is eluting during the first step gradient (compare to Fig. 2b and c), only tG of the first gradient was investigated. Further, flow rate and temperature also exhibited significant impact in the *in-silico* simulation and were therefore selected for the experimental robustness study. The influence of ion-pairing effects on selectivity and peak shape caused by certain amounts of TFA in the mobile phase cannot be predicted by the modeling software. Consequently, TFA concentration in mobile phase A and B was selected as fourth component for the experimental robustness study.

The four chosen factors (tG of the first gradient, flow rate, column temperature and TFA conc. in the mobile phase) were varied based on a response surface (central composite) design of experiment, as illustrated in Fig. 3a. The model was evaluated for linear, quadratic and interaction terms using the software MODDE. Centre-points were conducted as six-time repeated experiments. The response of the model was evaluated for the critical resolution between neighboring peaks. Factors were varied experimentally in the following ranges: duration of the first gradient from 7.5 to 8.5

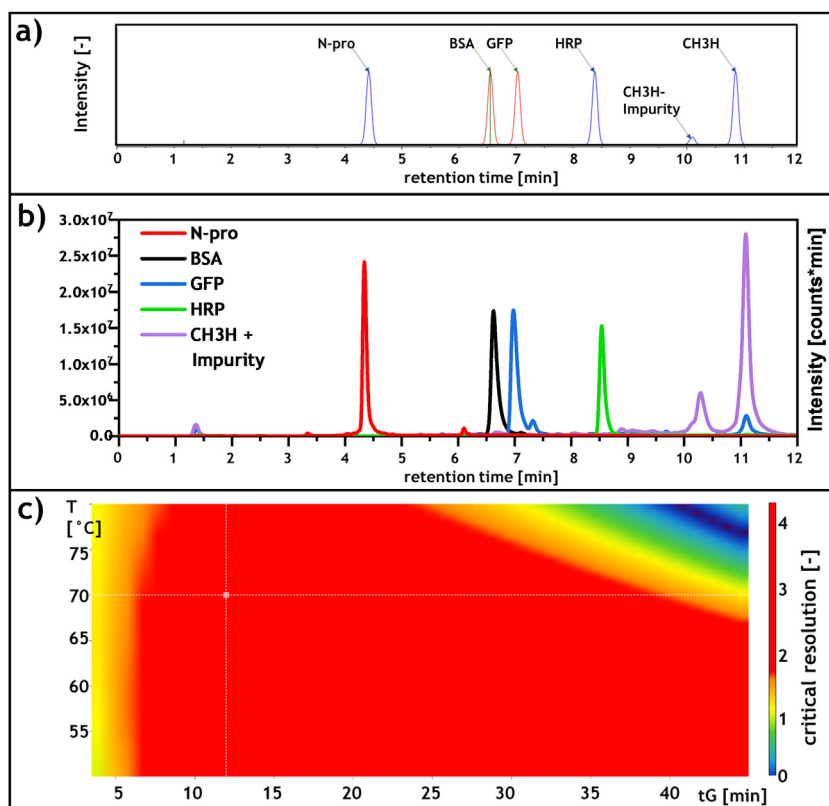


Fig. 2. a) Simulated *in-silico* chromatogram received from modeling with DryLab4; b) the chromatogram of the verification runs for all five target proteins utilizing the developed method; single chromatograms are depicted as overlay in order to show the baseline-separation of all applied target proteins c) heat map received from DryLab4 for chromatographic modeling for the corresponding input runs showing the influence of column temperature and gradient time.

Table 4

Statistical evaluation of the model for all resolutions performed via R^2 , R^2 adjusted and Q^2 .

	R^2	R^2 adj.	Q^2
Resolution.N-pro.BSA	0.962	0.937	0.778
Resolution.BSA.GFP	0.983	0.973	0.854
Resolution.GFP.HRP	0.941	0.924	0.811
Resolution.HRP.CH3H.Impurity	0.957	0.942	0.776
Resolution.CH3H.Impurity.CH3H	0.915	0.897	0.839

min, temperature from 67–73 °C, flow rate from 0.35–0.45 mL/min and TFA concentration from 0.08–0.12 %.

The model was evaluated for its statistical terms being the measure of fit (R^2) and the adjusted measure of fit (R^2 adj.) to validate the model's variation of response, adjusted for degrees of freedom. Furthermore, the model predictability Q^2 was evaluated to test for the model capability of predicting new data, shown in Table 4.

High R^2 -values (> 0.9) indicated that input data fit the model well. High values received for Q^2 (> 0.775) evaluated with MODDE showed that new data would fit the model properly and low experimental error was observed. To avoid any data specific model responses, the difference between R^2 and Q^2 should be as small as possible, which was found to be smaller than 0.184 for the stated data-set. The difference between R^2 and R^2 adj. indicates that the experimental setup performed within this study was chosen appropriate. Experiments conducted for all model proteins are given in Supplementary table 2B, whereas the resolution between the critical peak pair BSA-GFP will be discussed in more detail within this chapter, as these target proteins resulted in the smallest resolution (compare to Fig. 2b).

The DoE was evaluated by the resulting contour and interaction plots as illustrated in Fig. 3b and 3c. Even though column tempera-

tures below 70 °C were found to increase the critical resolution, 70 °C was considered as a compromise in order to avoid non-specific protein adsorption effects occurring at lower temperatures. Similar results were obtained for TFA concentration in the mobile phases: A lower TFA concentration was found to decrease the critical resolution, whereas a higher TFA concentration would improve critical resolution values but would on the other hand detrimentally impact sensitivity in mass spectrometry. A lower flow rate would improve the critical resolution but increase run time, simultaneously (see Fig. 3c). Therefore, a flow rate of 0.4 mL/min was determined as optimum. A short gradient time seemed to be beneficial, however the interaction plot in Fig. 3c showed that this was only true for smaller flow rates thus impeding a fast run time. Flow rates above 0.45 mL/min showed a greater critical resolution for long gradient times, compared to shorter gradient times.

No method factor combination resulted in a critical resolution lower than 1.5, hence selected optima were confirmed experimentally and the method can be considered robust for method variations such as ± 3 °C column temperature, ± 0.02 % of TFA concentration in mobile phase, ± 0.5 min of gradient time and ± 0.05 mL/min flow rate.

3.4. Method validation and comparison to a benchmark technology

3.4.1. Instrument precision and repeatability

Instrument precision was tested via six subsequent injections of the same protein solution at sample nominal concentration (compare to Section 2.3). A relative standard deviation (RSD) of at most 2.0 % for peak area and retention time were set as acceptance criteria for all five proteins. For instrument precision experiments conducted with BSA, an injection of a blank solution (7.5 M

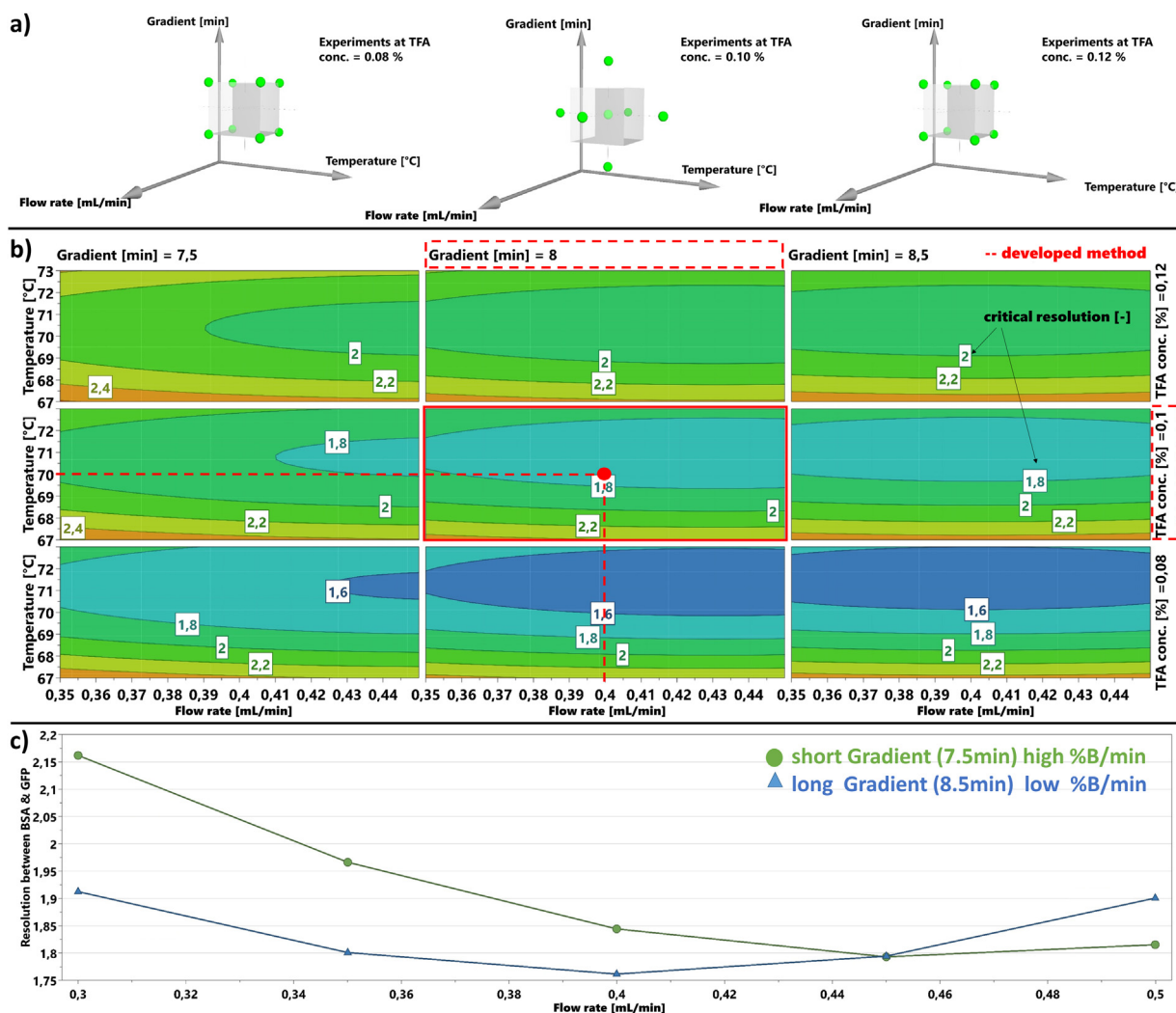


Fig. 3. a) Design of experiment and the experiments performed, exemplarily when keeping TFA conc. constant at 0.1 %; b) contour plot for all four factors indicating the best operating parameters to achieve a resolution greater than 1.5; c) an interaction plot, showing the dependencies of the variation of gradient and flow rate onto the critical resolution.

guanidine-HCl solution) was required to prevent sample carryover from one run to the next. RSD of the peak areas and the retention times were evaluated and fulfilled the criteria of not more than 2.0 % (compare to Supplementary Table C1).

In order to evaluate the impact of sample preparation, a repeatability study was performed. As BSA was commercially purchased, whereas other target proteins were derived from fermentation and subsequent purification, BSA was applied as model protein for the repeatability experiments. RSD for peak area and retention time below 2.0 % could be achieved throughout a six-fold sample preparation and subsequent analysis (compare to Supplementary Table C2).

3.4.2. Linearity and limit of quantitation

In biotechnology research and development model proteins are often quantified using BSA as a “generic” reference standard, in case no isolated material of the target protein is available [24]. For this reason, we applied BSA for evaluation of the Limit of Quantitation (LOQ) and linearity. A LOQ of 0.005 mg/mL could be established (RSD < 15 %, S/N ratio > 10). The method was found to be linear and precise over three orders of magnitude (correlation coefficient $R^2 = 0.999$, residual standard deviation = 1.7 %, prognosis interval $P = 95$ % for the y-intercept includes the zero point). The calibration curve

is shown in Fig. 4a and further details on linearity evaluation are given Table 5.

3.4.3. Comparison of the developed RPLC method with SDS-PAGE analysis

Even though analysis of microbial proteins is commonly performed with SDS-PAGE, proper LC analytics should be superior regarding accuracy and analysis time [5]. To verify this hypothesis, a standard calibration curve established with (i) SDS-PAGE analysis (using comassie blue staining) was compared to (ii) a calibration curve measured with the herein developed RPLC method. Method attributes such as analysis time, LOQ, correlation coefficient, y-intercept and RSD are compared in Table 5.

The standard calibration curves (compare to Table 5 and Fig. 4a and b) show that higher accuracy and a decrease in analysis time can be achieved using the established RPLC method. Due to method limitations low protein concentrations could not be evaluated with a gel-based quantification, as indicated in Fig. 4b. As integration of SDS-PAGE-derived bands is challenging and needs to be performed visually (being highly operator dependent), high deviations in evaluation can be observed, which is in accordance with literature [6]. To test whether host cell impurities would impact the resolution of target proteins, crude samples were compared with purified ones (see Supplementary data section D, figure D1). The respective target

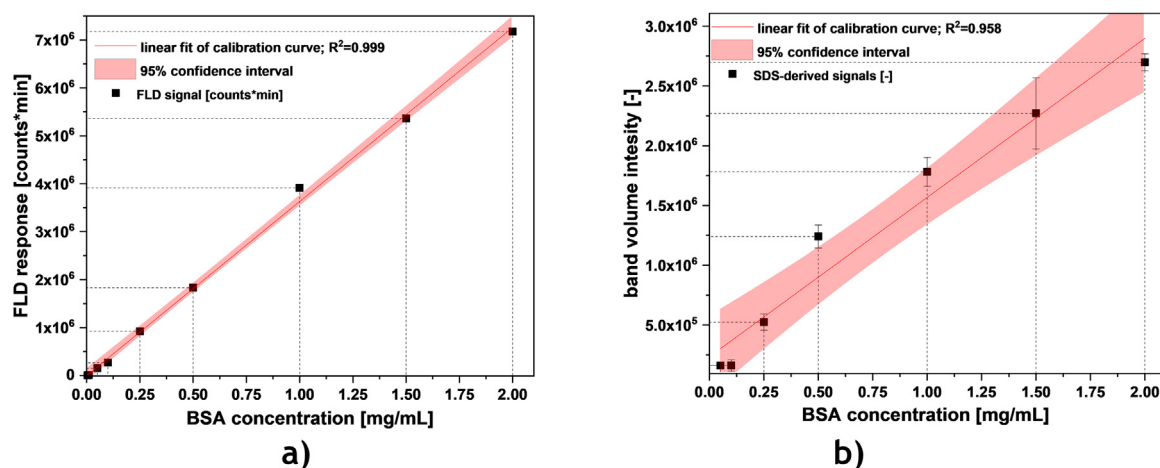


Fig. 4. Calibration of BSA standards with a confidence interval set for 95 % comparing a) the developed RPLC method with b) a state of the art SDS-PAGE using comassie blue staining.

Table 5

Parameters for the evaluation of the two compared analytical methods, using LOQ, correlation coefficient, y-intercept, and residual standard deviation referred to nominal concentration.

Method	Analysis time [min/sample]	LOQ BSA [mg/mL]	correlation coefficient [-]	y-intercept (ref. to sample nominal concentration) [%]	residual standard deviation (ref. to sample nominal concentration) [%]
RPLC	18	0.005	0.999	0	1.7
SDS-PAGE	150	0.1	0.979	8.7	8.6

peaks could be separated from their impurities, thus eliminating the need for SDS-page derived analysis throughout any process step.

Summarizing the comparison of both methods, RPLC measurements outperform the SDS-PAGE in terms of analysis speed (18 min versus 150 min) and limit of quantification (LOQ 0.005 mg/mL versus 0.1 mg/mL). This makes RPLC an attractive technique facilitating pharmaceutical up- and downstream development where fast and accurate responses are a prerequisite for screening conditions that potentially increase recovery yields and/or purity of the final product.

4. Conclusion

A generic RPLC method for quantification of commonly applied model proteins expressed by the microbial host *E. coli* was developed. The method development workflow applied analytical quality-by-design principles by using retention modeling and establishing a method operable design region. The MODR was challenged by both *in-silico* and experimental robustness experiments, using a multivariate experimental design and analysis approach. Method performance of the established RPLC method was compared to a SDS-PAGE method, which is still a benchmark technology for the analysis of microbial derived samples. Results indicate the superiority of the RPLC method regarding analysis time, ease of use, linear range and LOQ.

In addition, the herein described AQbD workflow can be used as a generic method development protocol for RPLC analysis of proteins. Furthermore, the presented method has found direct utilization in the analysis of the chosen target proteins, indicating (i) its applicability for the analysis of a broad range of proteins with different hydrophobicities and (ii) its independency of applied protein purification procedures and the therefore needed buffer solutions. Finally, the established RPLC method is also MS-compatible thus enabling further opportunities for mass elucidation.

Author contributions

JK and AP performed analytical experiments. FZ constructed the experimental robustness study. RP conducted DryLab experiments. JH, DH and JE produced and purified proteins and performed experiments. JK, FZ and RP evaluated experiments and performed data treatment. CH, OS and CS gave valuable scientific input. JK, FZ, CS and RP drafted the manuscript.

Funding

The authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program. Furthermore, the authors thank the Christian Doppler Society for funding of this project.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jpba.2020.113412>.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Furthermore, the authors declare that they have no competing interests.

References

- [1] G. Walsh, Biopharmaceutical benchmarks 2018, *Nat. Biotechnol.* 36 (12) (2018) 1136–1145.
- [2] J. Kopp, C. Slouka, O. Spadiut, C. Herwig, The rocky road from Fed-Batch to continuous processing with *E. coli*, *Front. Bioeng. Biotechnol.* 7 (2019) 328.
- [3] A.S. Rathore, QbD/PAT for bioprocessing: moving from theory to implementation, *Curr. Opin. Chem. Eng.* 6 (2014) 1–8.

- [4] M.H. Caparon, K.J. Rust, A.K. Hunter, J.K. McLaughlin, K.E. Thomas, J.T. Herberg, R.E. Shell, P.B. Lanter, B.F. Bishop, R.L. Dufield, Integrated solution to purification challenges in the manufacture of a soluble recombinant protein in *E. coli*, *Biotechnol. Bioeng.* 105 (2) (2010) 239–249.
- [5] J. Sasse, S.R. Gallagher, Staining proteins in gels, *Curr. Protoc. Mol. Biol.* 85 (1) (2009), 10.6.1–10.6.27.
- [6] V. Dolnik, W.A. Gurske, Chemical modification of proteins to improve the accuracy of their relative molecular mass determination by electrophoresis, *Electrophoresis* 32 (20) (2011) 2893–2897.
- [7] E. Domínguez-Vega, R. Haselberg, G.W. Somsen, Capillary zone electrophoresis–Mass spectrometry of intact proteins, in: N.T. Tran, M. Taverna (Eds.), *Capillary Electrophoresis of Proteins and Peptides: Methods and Protocols*, Springer, New York, New York, NY, 2016, pp. 25–41.
- [8] S. Fekete, D. Guilleme, P. Sandra, K. Sandra, Chromatographic, electrophoretic, and mass spectrometric methods for the analytical characterization of protein biopharmaceuticals, *Anal. Chem.* 88 (1) (2016) 480–507.
- [9] R.L.C. Voeten, I.K. Ventouri, R. Haselberg, G.W. Somsen, Capillary electrophoresis: trends and recent advances, *Anal. Chem.* 90 (3) (2018) 1464–1481.
- [10] K. Sandra, I. Vandenheede, P. Sandra, Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization, *J. Chromatogr. A* 1335 (2014) 81–103.
- [11] S. Fekete, A. Beck, J.L. Veuthey, D. Guilleme, Ion-exchange chromatography for the characterization of biopharmaceuticals, *J. Pharm. Biomed. Anal.* 113 (2015) 43–55.
- [12] S. Fekete, D. Guilleme, Ultra-high-performance liquid chromatography for the characterization of therapeutic proteins, *Trac Trends Anal. Chem.* 63 (2014) 76–84.
- [13] S. Fekete, J.-L. Veuthey, D. Guilleme, Modern Column Technologies for the Analytical Characterization of Biopharmaceuticals in Various Liquid Chromatographic Modes, LCGC Europe, 2015.
- [14] M. Baca, J. De Vos, G. Bruylants, K. Bartik, X. Liu, K. Cook, S. Eeltink, A comprehensive study to protein retention in hydrophobic interaction chromatography, *J. Chromatogr. B* 1032 (2016) 182–188.
- [15] S. Fekete, A. Goyon, J.-L. Veuthey, D. Guilleme, Size exclusion chromatography of protein biopharmaceuticals: past, present and future, *Am. Pharm. Rev.* (2018) 1099–8012.
- [16] T. Janaratne, X.C. Wang, C. Becker, Y. Zhao, R. Leanna, W. Pritts, Hydrophobic interaction chromatography (HIC) method development and characterization of resolved drug-load variants in site-specifically conjugated pyrrolbenzodiazepine dimer-based antibody drug conjugates (PBD-ADCs), *J. Pharm. Biomed. Anal.* 179 (2020) 113027.
- [17] S. Fekete, J.-L. Veuthey, D. Guilleme, New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: theory and applications, *J. Pharm. Biomed. Anal.* 69 (2012) 9–27.
- [18] J. Kochling, W. Wu, Y. Hua, Q. Guan, J. Castaneda-Merced, A platform analytical quality by design (AQbD) approach for multiple UHPLC–UV and UHPLC–MS methods development for protein analysis, *J. Pharm. Biomed. Anal.* 125 (2016) 130–139.
- [19] B. Bobály, V. D'Atri, M. Lauber, A. Beck, D. Guilleme, S. Fekete, Characterizing various monoclonal antibodies with milder reversed phase chromatography conditions, *J. Chromatogr. B* 1096 (2018) 1–10.
- [20] B. Bobály, J.-L. Veuthey, D. Guilleme, S. Fekete, New developments and possibilities of wide-pore superficially porous particle technology applied for the liquid chromatographic analysis of therapeutic proteins, *J. Pharm. Biomed. Anal.* 158 (2018) 225–235.
- [21] B. Bobály, M. Lauber, A. Beck, D. Guilleme, S. Fekete, Utility of a high coverage phenyl-bonding and wide-pore superficially porous particle for the analysis of monoclonal antibodies and related products, *J. Chromatogr. A* 1549 (2018) 63–76.
- [22] B. Sylvestre, L. Tefas, L. Vlase, I. Tomuță, A. Porfire, A Quality by Design (QbD) approach to the development of a gradient high-performance liquid chromatography for the simultaneous assay of curcuminoids and doxorubicin from long-circulating liposomes, *J. Pharm. Biomed. Anal.* 158 (2018) 395–404.
- [23] A. Tumpa, A. Stajić, B. Jančić-Stojanović, M. Medenica, Quality by Design in the development of hydrophilic interaction liquid chromatography method with gradient elution for the analysis of olanzapine, *J. Pharm. Biomed. Anal.* 134 (2017) 18–26.
- [24] D.J. Wurm, J. Quehenberger, J. Mildner, B. Eggenreich, C. Slouka, A. Schwaighofer, K. Wieland, B. Lendl, V. Rajamanickam, C. Herwig, O. Spadiut, Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in *E. coli*, *Appl. Microbiol. Biotechnol.* (2017).
- [25] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, the Proteomics Protocols Handbook, Springer, 2005, pp. 571–607.
- [26] F. Ferré, P. Clote, DiANNA: a web server for disulfide connectivity prediction, *Nucleic Acids Res.* 33 (suppl.2) (2005) W230–W232.
- [27] S. Fekete, V. Sadat-Noorbakhsh, C. Schelling, I. Molnar, D. Guilleme, S. Rudaz, J.L. Veuthey, Implementation of a generic liquid chromatographic method development workflow: application to the analysis of phytocannabinoids and Cannabis sativa extracts, *J. Pharm. Biomed. Anal.* 155 (2018) 116–124.
- [28] R.M. Krisko, K. McLaughlin, M.J. Koenigbauer, C.E. Lunte, Application of a column selection system and DryLab software for high-performance liquid chromatography method development, *J. Chromatogr. A* 1122 (1) (2006) 186–193.
- [29] I. Molnar, Computerized design of separation strategies by reversed-phase liquid chromatography: development of DryLab software, *J. Chromatogr. A* 965 (1) (2002) 175–194.
- [30] B. Dejaegher, Y.V. Heyden, Ruggedness and robustness testing, *J. Chromatogr. A* 1158 (1) (2007) 138–157.