



Review article

Advances in downstream processing of biologics – Spectroscopy: An emerging process analytical technology



Matthias Rüdert, Till Briskot, Jürgen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte Ring 3, 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:

Received 11 October 2016

Received in revised form 7 November 2016

Accepted 8 November 2016

Available online 15 November 2016

Keywords:

Downstream processing

Process analytical technology

Spectroscopy

Chemometrics

Biologics

ABSTRACT

Process analytical technologies (PAT) for the manufacturing of biologics have drawn increased interest in the last decade. Besides being encouraged by the Food and Drug Administration's (FDA's) PAT initiative, PAT promises to improve process understanding, reduce overall production costs and help to implement continuous manufacturing. This article focuses on spectroscopic tools for PAT in downstream processing (DSP). Recent advances and future perspectives will be reviewed. In order to exploit the full potential of gathered data, chemometric tools are widely used for the evaluation of complex spectroscopic information. Thus, an introduction into the field will be given.

© 2016 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/>).

Contents

1. Introduction	2
2. Multivariate data analysis for PAT	3
2.1. Multivariate projection methods	3
2.2. Principal component analysis	3
2.3. Partial least square regression	4
3. Spectroscopy for process monitoring in chromatography	4
3.1. UV/vis spectroscopy	6
3.2. FTIR spectroscopy	6
3.3. Other spectroscopic PAT tools	7
4. Conclusion and outlook	7
Acknowledgements	7
References	7

1. Introduction

In 2004, the United States' FDA published *Guidance for industry. PAT – A framework for innovative pharmaceutical development, manufacturing and quality assurance* [1]. Within the guidance, FDA promotes the implementation of PAT into all unit operations to monitor critical quality attributes (CQAs). PAT is described as being part of process design and furthermore intended to contribute

to process control, i.e. to be taken actively into account for process decisions. While being intended for both small molecules and biologics, the implementation into these two domains of pharmaceuticals is advancing at different paces. In the past, PAT was adopted more quickly in the production of small molecules. For an extensive review thereof, the authors defer to [2]. This article will focus on biologics only.

In contrast to most small molecules, biologics are produced in living organisms which are very sensitive to a wide variety of external factors. Most biologics are complex proteins. They do not consist of one chemical entity but a distribution of many species. Already slight process changes can affect the product quality profile [3]. In order to ensure a consistent product quality and to reduce

* Corresponding author.

E-mail address: juergen.hubbuch@kit.edu (J. Hubbuch).

batch-to-batch variability, PAT for biologics is of great interest [4–8]. Furthermore, current trends towards continuous manufacturing may require an improved process control for keeping a steady state over prolonged periods of time. Such a control may be simplified by the possibility to monitor critical quality attributes in real-time [9].

Other advantages of PAT include the simplification of root cause analysis [10] and improvement of process understanding. Eventually, the improved process understanding and real-time monitoring capabilities may lead to the implementation of the concept of real-time release [1]. Thus, the CQA profile of the final product can be guaranteed to lie within acceptable quality limits solely based on real-time measurements and production batches can be released based on this data.

Early approaches to PAT for biologics widely addressed the problem by implementing on-line analytical chromatography. Already before the release of the PAT guidance, on-line high performance liquid chromatography (HPLC) has been used to control column loading and pooling decisions during chromatographic purification steps [11,12]. Subsequently, on-line and at-line HPLC was further used for a variety of applications [13–15]. Recently, at-line HPLC has been also implemented in the control of continuous chromatography equipment [16]. HPLC provides high resolution of different species. However, it is complex regarding the required equipment, consisting of a device for sampling as well as the chromatograph itself. This may be undesirable in a manufacturing environment as reliability may be an issue. Furthermore, automated sampling and the analytical separation also lead to non-negligible time delays. Depending on the decision time of a unit operation, this may lead to late notice of process deviations or even completely prevent real-time monitoring.

Spectroscopy is a powerful tool for process monitoring [17]. Spectroscopic equipment has similar investment costs (\$20k to \$200k) as on-line HPLC. Measurement times are fast, typically in the subsecond range up to a few minutes. Furthermore, measurements can often readily be performed in-line. Fast measurement times are especially important for preparative chromatography, the workhorse in current DSP. Preparative chromatographic processes are highly non-linear and feature sharp concentration fronts [18]. Thus, CQAs of the effluent such as the mass fraction of impurities are quickly changing. To reliably control such processes, the used monitoring method needs to have short response times. Typical decision times for preparative protein chromatography lie in the range of 30 s to several minutes. In contrast to at-line HPLC, spectroscopy provides signals with limited selectivity for different components. To overcome this limitation, a combination of multivariate measurements and mathematical tools for multivariate data analysis (MVDA) is generally applied to extract information from spectroscopic measurements.

Following this argumentation, this article is focusing in a first part on the review of two widely used chemometric tools for the analysis of spectroscopic data. Subsequently, the current state-of-art of spectroscopic PAT in DSP is discussed.

2. Multivariate data analysis for PAT

The implementation of the PAT framework is often accompanied by the application of *multivariate mathematical approaches* [1], also known as chemometrics. In chemometrics, mathematical and statistical tools are used to extract useful chemical information from large amounts of multivariate measurements or raw data [19]. The multivariate nature of spectroscopic data for PAT arises out of necessity, since no univariate process analyzer has significant selectivity to monitor a specific CQA without interferences from other properties [17]. Chemometrics can be used for a wide

variety of tasks, including experimental design (DoE) and MVDA [20]. The present article does not aim to give a complete review of all elements in chemometrics, but focuses solely on MVDA. Furthermore, only the two most common MVDA tools in PAT are discussed more closely: principal component analysis (PCA) and partial least squares regression (PLS). A more thorough review of chemometric tools is given in the textbook of Bakeev [17].

2.1. Multivariate projection methods

Multivariate projection (decomposition) to latent structures forms the basis of many approaches in MVDA [21]. According to Kvalheim [22,23], the latent variable (LV) projection of a data matrix $X = (\tilde{x}_1, \dots, \tilde{x}_k)$, with n observations and k variables, can most easily be understood by reference to variable and objective space, as illustrated in Fig. 1. The former case (Fig. 1a) reveals relationships between observations by plotting the observations in a space spanned by the k variables in X . In the object space (Fig. 1b), the coordinate system is defined by the n observations. It visualizes information about the relationship between variables [22]. The main goal of latent projection methods is to reduce the dimensions in the variable space by summarizing variables with similar information in LVs. All latent projection methods help getting fundamental insights into complex multivariate data by (1) discovering groupings in the data, (2) data compression, (3) regression, and more [24].

The variable decomposition into LVs can geometrically be interpreted as a projection of the data in the variable and object space on a -dimensional hyperplanes, whereby a represents the number of LVs. Since the projection is performed in both spaces, the maximum number of LVs is $\min(n, k)$. The projection coordinates (scores) of the observations in the variable space on the i th LV are summarized in the score vector $\tilde{t}_i$ and are obtained by projecting the samples on the corresponding weight vector $\tilde{w}_i$ [23]. The vectors $\tilde{t}_i$ and $\tilde{w}_i$ are orthogonal and orthonormal, respectively. Any latent projection method can be derived over the definition of $\tilde{w}_i$ [20]. The projection coordinates (loadings) of the variables in the object space are summarized in the loading vector $\tilde{p}_i$. The loading vectors $\tilde{p}_i$ are not necessarily orthogonal.

2.2. Principal component analysis

PCA is a common tool in exploratory data analysis and is used for data reduction, simplification, outlier detection, classification, and noise reduction [25]. Data decomposition of a matrix X according to

$$X = TP^T + E_X \quad (1)$$

is performed with the objective to explain as much as of the variance in X by a linear combination of a complementary set of scores $T = (\tilde{t}_1, \dots, \tilde{t}_a)$ and loadings $P = (\tilde{p}_1, \dots, \tilde{p}_a)$. In order to differentiate the data decomposition by PCA from other latent projection methods, the LVs are referred to as principal components (PCs). In PCA, the loadings $\tilde{p}_i$ are equal to the weights $\tilde{w}_i$ and thus orthonormal. They give a quantitative measure of the part of variance and observed variable shares with the PC [22]. Thus, the whole information regarding the linear relationship between variables is compressed in the loading matrix P . The hidden structure of X concerning the object space can be visualized by loading plots, where the loadings $\tilde{p}_i$ are plotted against each other [25]. Variables having similar loading values on a PC are linear dependent (collinear) and are redundant concerning this PC. For mean centered data, as illustrated in Fig. 1, collinearity between two variables can graphically be visualized by the cosine of the angle between the two variables in the object space. In the same manner as relationships between variables can be illustrated by loading plots, relationships

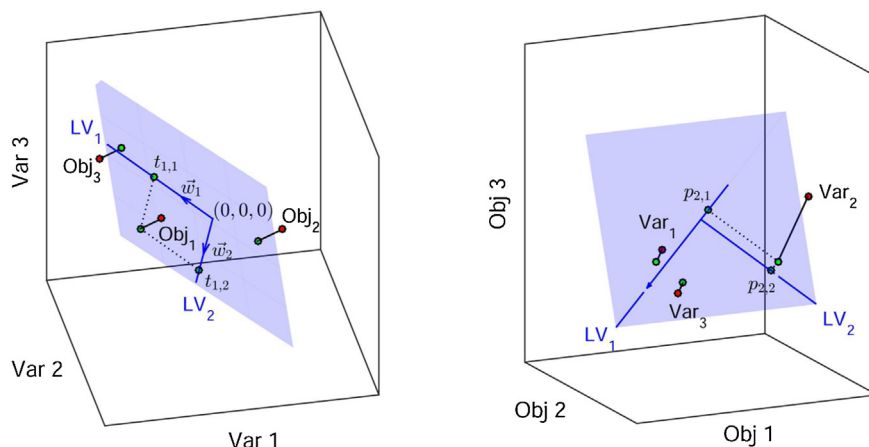


Fig. 1. Visualization of a data matrix X consisting of three observations with three variables in the variable space (a) and object space (b). Observations in the variable space are projected on a latent structure defined by the weight vectors $\bar{w}_i$, leading to the projection coordinates (scores) $\bar{t}_i$. Projection coordinates (loadings) of the variables in the object space are summarized in the loading vectors $\bar{p}_i$.

between observations can be visualized by score plots [25]. Score plots can reveal patterns, clusters, and outliers in the observations (measurements). Usually, two or three PCs are already sufficient to reveal hidden patterns in X by loading and score plots, since the most useful information (variance) in X is explained by the first few PCs. The remaining ones are assumed to comprise predominantly noise [25]. By neglecting these minor PCs, PCA achieves a data simplification and noise reduction in X . Since both scores and loadings are orthogonal, PCA is also able to reduce collinearity in X , which is why it also plays a central role in regression analysis.

2.3. Partial least square regression

Linear regression methods like PLS are tools in exploratory data analysis, relating one or more response variables Y with several predictor variables X , by a linear multivariate model

$$Y = XB + E_Y \quad (2)$$

where B contains the regression coefficients connecting the predictor variables to the responses. The deviation between model responses and measurements is summarized in the residual matrix E_Y . In the simplest case, when the matrix X is of full rank, multiple linear regression (MLR) can be applied and the regression coefficients can be obtained by the least square solution

$$B = (X^T X)^{-1} X^T Y. \quad (3)$$

In most PAT applications, however, the observation to variable ratio is rather low and the X -variables are collinear and noisy. In such cases, prediction abilities of MLR models can be very poor since the estimated regression coefficients become unstable and can deviate substantially from their expected values [26,27].

An alternative way to determine the regression coefficients B is by using latent projection methods like principal component regression (PCR) and PLS. In PCR the collinearity problem is solved by (1) decomposing the predictor matrix X to orthogonal PCs and (2) regressing the responses Y on the orthogonal scores T instead of X . The score matrix T is of full rank and allows the prediction of stable regression coefficients. Furthermore, data decomposition prior to regression allows noise reduction and thus the calibration of more robust models. A major drawback of PCR is that data decomposition is performed under the objective to explain as much as possible of the variance in X . However, the variance in X that is relevant for the prediction of Y could be rather small in comparison with the total variance in X . Thus, much of the relevant variance could be lost by PCA [17].

In contrast to PCR, PLS performs a simultaneous decomposition of X and Y with the objective to explain as much as possible of the covariance between the data sets [28]. The decomposition of X and Y can be described by

$$T = XW \quad (4)$$

and

$$Y = UC^T + E_C \quad (5)$$

where $U = (\bar{u}_1, \dots, \bar{u}_a)$ contains the corresponding Y -scores $\bar{u}_i$ on the i th latent variable, E_C represents the Y -residuals, and $C = (\bar{c}_1, \dots, \bar{c}_a)$ denotes the linear transformation defined by the orthogonal Y -loadings $\bar{c}_i$. Since the weight matrix W is determined under the objective of maximizing covariance between X and Y , the scores T are good predictors of the original data X

$$X = TP^T + E_X \quad (6)$$

and model also the responses [29]

$$Y = TC^T + E_Y. \quad (7)$$

In contrast to PCA, weights $\bar{w}_i$ and loadings $\bar{p}_i$ are not equal. The orthonormal weights can be considered as *tilted* X -loadings since they describe the effective relationship between X and Y . Depending on how strong Y effects W , the weights $\bar{w}_i$ deviate more or less from the loadings $\bar{p}_i$ [30]. The X -loadings are not orthogonal to each other [24]. Comparing Eq. (2) with Eq. (7) leads to the regression coefficient

$$B = WC^T. \quad (8)$$

Since the regression model B is calculated from the orthogonal latent structures W and C , PLS is able to analyze data with strongly collinear, noisy, and numerous X -variables [29].

3. Spectroscopy for process monitoring in chromatography

In the past, spectroscopic methods have been widely used as tools for structural analysis of proteins [31–33]. From a biochemical point of view the analysis of proteins can be split into the assessment of primary, secondary, tertiary and quaternary structures. Spectroscopic methods provide information about each of these layers of abstraction within the protein structure (cf. Fig. 2) [31]. To assess the sequence and total concentration of protein, especially UV/vis spectroscopy and Fourier-Transform Infrared (FTIR) spectroscopy are of interest. UV/vis spectroscopy mainly measures the primary structure, i.e. the content of aromatic amino acids as

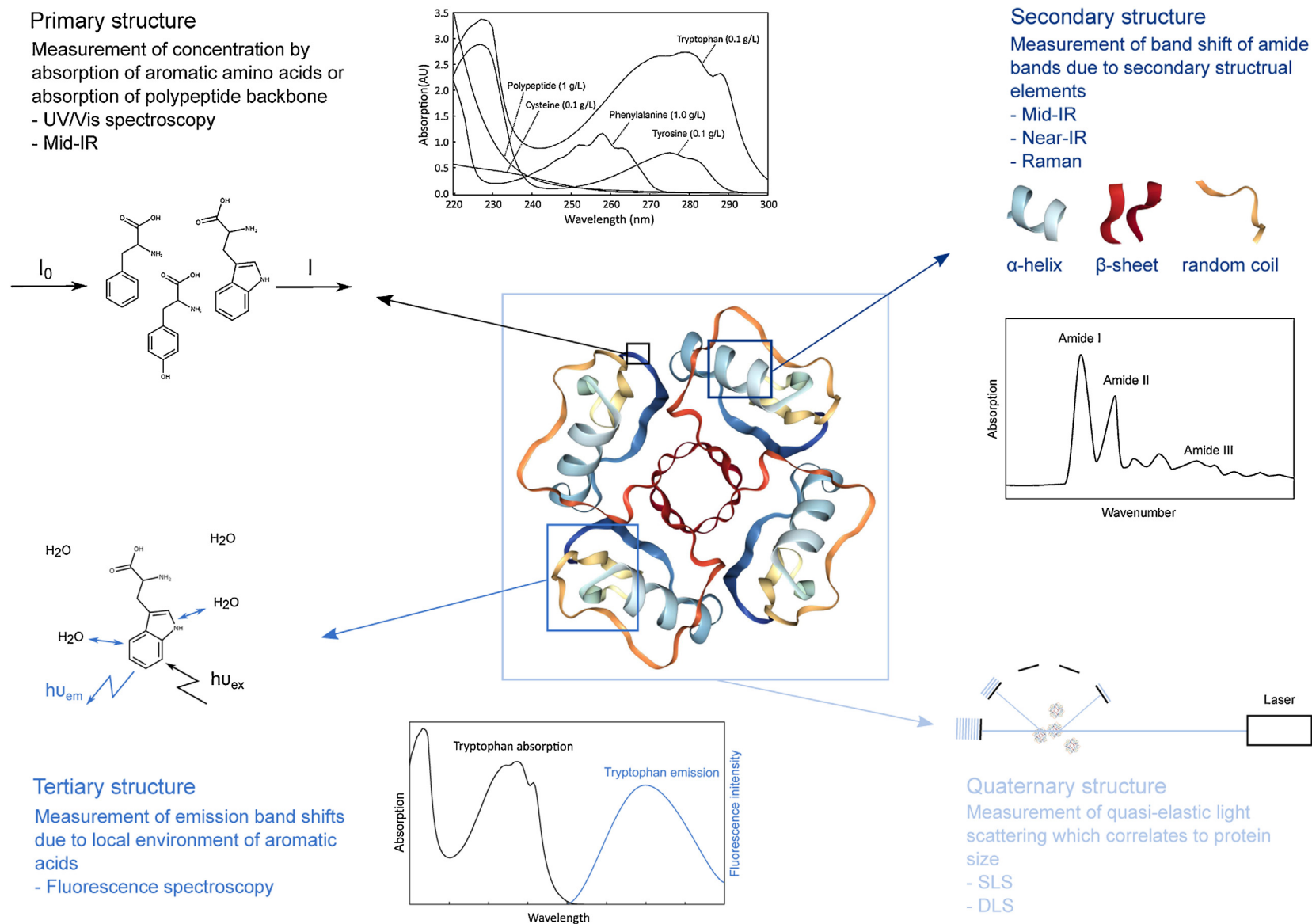


Fig. 2. Based on the example of ovomucoid, the four different level of protein structure are illustrated. To each level, suitable spectroscopic methods are listed with a short explanation of what is measured. The lists are not extensive but rather correspond to the most promising methods in the authors eyes. Protein structure retrieved from PDB ID: 1OVO [53,54]. UV/vis spectra obtained from [36].

well as weak spectral shifts due to the solvochromatic effects [31]. The secondary structure of proteins has been frequently measured by vibrational spectroscopy such as FTIR and Raman spectroscopy [32,34,35]. The methods allow to measure the vibrational modes of the backbone of polypeptides. The tertiary structure of proteins is accessible over the fluorescence of the aromatic amino acids. The tryptophan emission is solvatochromatic, reacting to changes in the local polarity around tryptophan residues [31,33]. Thus, structural changes which affect the local environment of tryptophan residues can be detected by fluorescence spectroscopy. Finally, the quaternary structure of proteins, i.e. assembly of multiple subunits or native aggregation of protein monomers, may be assessed over the protein size by quasi-elastic light scattering methods including static light scattering (SLS) and dynamic light scattering (DLS) [31,4].

All of the above mentioned methods are of major interest for process monitoring as each method provides access to orthogonal information about the product. Key aspects of the different methods have been summarized in Table 1. In literature, especially UV/vis absorption and FTIR have been used for a variety of PAT applications (cf. Sections 3.1 and 3.2). Literature for fluorescence spectroscopy as well as DLS is less broad. However interesting applications exist (cf. Section 3.3). In the following sections, the different applications will be discussed.

3.1. UV/vis spectroscopy

UV/vis spectroscopy measures the absorption of proteins generally in the range between 240 and 340 nm. Mainly due to the content of aromatic amino acids (phenylalanine, tyrosine, and tryptophan) proteins significantly absorb in this region (cf. Fig. 2, primary structure) [18,31,36]. Due to the sensitivity, reproducibility of signals and robustness of the spectrometers, UV/vis absorption at 280 nm is widely used as a primary detection method of protein concentrations. While current applications mainly rely on univariate UV/vis measurements, it has been shown that UV/vis spectra contain a significant amount of information on proteins and may be used for selective quantification even if only minute spectral differences exist [36].

Multivariate UV/vis spectroscopy in conjunction with PLS modeling for selective protein quantification first appeared in 1994 [37]. Arteaga et al. demonstrated the quantification of the three main bovine caseins by PLS regression on the fourth derivative UV/vis spectra. The PLS model was calibrated based on designed mixing ratios. In contrast to latter publications which focus on (near)-real-time assays, Arteaga et al. intended the proposed method as an off-line analytical assay. In the scope of the publication, the method was not applied to process samples.

The first at-line application for chromatography was only reported in 2011 as a tool to circumvent the analytical bottleneck created by high throughput experimentation [38]. Similar to Arteaga et al., a PLS model was calibrated based on designed mixing ratios of pure protein components. The calibrated PLS model was used to selectively quantify the protein content in elution fractions of multiple co-eluting species from miniaturized and parallelized chromatography experiments. The results were later confirmed by [39]. Subsequently, the method was transferred to an in-line setup with a diode array detector and applied for a selective and real-time quantification of 3 model proteins [40]. It was shown that the deconvoluted signal from the detector could be directly used in a feed-forward controller to trigger product pooling. Experiments were performed in diluted conditions to prevent detector saturation.

While the above mentioned publications provided accurate predictions of protein concentrations in multi-protein mixtures, they all relied on designed mixing ratios of pure proteins. This may

pose major difficulties when calibrating a PLS model for an applied example, e.g. the purification of a monoclonal Antibody (mAb) from its high molecular weight impurities (HMWs). Brestrich et al. addressed this problem by using process based samples for the PLS model calibration [41]. Instead of using pure protein samples to produce designed mixing ratios, chromatographic runs at variable conditions were performed to span a model calibration space. The column effluent of those experiments was fractionated and analyzed by suitable off-line analytics. They applied the newly designed method to different diluted separation problems including the separation of a mAb from its impurities and the measurement of different protein species in human blood fractionation.

Since then, UV/vis spectroscopy in conjunction with PLS modeling has been used in multiple studies. As a supportive tool, it was applied together with mechanistic modeling for a generic root-cause investigation [10]. In a first preparative in-line application, the tool has been used to monitor and control a chromatographic Protein A capture step, an application which may be of interest for controlling continuous chromatography equipment [42]. Current research aims to extend the applicable concentration range for the approach. During the elution of preparative chromatographic steps, very high protein concentrations may occur and cause detector saturation in the UV/vis range. By applying variable pathlength spectroscopy, the linear range of UV/vis absorption spectroscopy can be greatly extended. PLS modelling allowed the deconvolution of co-eluting species in multiple case studies [43].

3.2. FTIR spectroscopy

FTIR spectroscopy is frequently applied as a PAT technology for small molecule production [2]. For proteins, FTIR was first established as a tool for assessing the secondary structure [31,32,34,35]. Proteins are detected by the vibration of the polypeptide backbone. Multiple vibrational modes correspond to different detected amide bands (cf. Fig. 2, primary and secondary structure). The absorption of the amide bands is directly proportional to the amount of polypeptide backbone. The most prominent proteinogenic band, the amide I band, is mainly caused by C=O stretching. Secondary structural elements induce band shifts of the amide bands. This phenomenon can be used to quantify the proportion of different secondary structural elements, e.g. by taking the second derivative or applying Fourier self-deconvolution. Thus, FTIR is a promising candidate for monitoring the overall protein mass as well as the structural integrity of proteins by their secondary structure. The application is however hindered by the strong absorption of water in the same spectral region. It is a non-trivial task to correct for the water absorption. To prevent total extinction in the transmission cell, typical pathlengths need to be very short (approximately 5 μm), which however also reduces the sensitivity towards proteins. Despite the existing problems, a number of promising applications have been reported.

Publications demonstrated the possibility to selectively detect mAbs, HMWs and host cell proteins (HCPs) [44,45] with FTIR for biopharmaceutical applications. Experiments were performed in an at-line setup. The approach was later extended to further downstream processing unit operations [46]. Capito et al. demonstrated the use of a calibrated PLS model to selectively quantify mAb, HMW and HCP concentrations of samples drawn from different unit operations. Again experiments were performed at-line. mAb could be quantified down to concentrations of 0.7 g/l while HMW concentrations as low as 1% [w/w] were detected.

During the refolding process of an inclusion body of an auto-protease, FTIR was applied as an in-line PAT tool to monitor the relative content of different secondary structural elements [47]. A time evolution of the relative content of structural elements could be shown during the refolding process. However, the results did

Table 1

Summary of different spectroscopic methods of interest with key information on each method.

Spectroscopic method	Wavelength range	Acquisition time	Signal-to-noise ratio	Measured attribute	Remarks	References
UV/vis spectroscopy	240–340 nm	0.01–30 s	+	Mainly aromatic amino acids	Sensitive, quantitative, simple instrumentation	[10,31,36–38,40–43]
Mid-IR	2.5–25 μ m	Single scan 0.5–4 s	--	Peptide backbone	Strong water bands, for high signal-to-noise ratios multiple scans (100–600) are necessary	[32,34,45,44,46–48]
Near-IR	0.8–2.5 μ m	Single scan 0.5–4 s	–	Vibrational overtones of peptide backbone	Strong water bands, low sensitivity, low selectivity, simple instrumentation, fiber probes readily available	[17]
Raman spectroscopy	Depends on excitation laser	Depends on excitation laser	--	Peptide backbone	Generally low sensitivity, high selectivity, not infringed by water absorption, fiber probes readily available	[31,34]
Fluorescence spectroscopy	Excitation: 240–300 nm Emission: 260–450 nm	0.01–300 s	++	Aromatic amino acids and their solvchromatic environment	Broad measurement ranges feasible, calibration may be challenging	[33,31,49]
DLS	Visible light, e.g. 633 nm	0.5–8 min	–	Diffusion behaviour of macromolecules	Based on time correlation, limited suitability for flow-through	[50]
SLS	Visible light, e.g. 633 nm	<1 s	–	Light scattering due to particle size	Difficult to obtain stable baselines	[31]
Circular dichroism	190–260 nm	<1 s	+	Chirality of secondary structure	Impractical at high concentrations	[31]

not allow prediction of the refolding yield based on the computed content of secondary structural elements.

Recently, an approach was published to monitor the in-column binding behaviour of mAb during a Protein A capture step by FTIR [48]. A micro-column was packed on top of an attenuated total reflection (ATR) crystal. With a PLS model, the total protein content of resin in contact with the ATR crystal was measured over multiple process steps. The publication showed, that the clean-in-place steps do not seem to be able to reliably remove all bound protein. While being an interesting scientific approach, a transfer to a larger scale system may be difficult. The proposed setup samples the resin very locally, which may not be representative for the overall column. Furthermore, lateral stress had to be applied to generate an increased contact area between resin and ATR crystal. Nevertheless, the approach shows the versatility of FTIR spectroscopy.

3.3. Other spectroscopic PAT tools

To the best of our knowledge, other spectroscopic methods have only been studied by two articles as PAT technologies for DSP of biologics. Fluorescence spectroscopy was proposed as an at-line PAT tool for a chromatographic purification step of a fusion protein [49]. Here, it was shown that the fluorescence signal could be correlated with the fraction purity from an hydrophobic interaction chromatography step separating misfolds from the product. DLS was used to investigate the unfolding and refolding process of a recombinant fusion protein from an inclusion body and its dependence on a chaotropic agent [50]. Yu et al. could accurately predict the aggregation and folding state compared to reference analytics. The method was however not applied for real-time process monitoring or control.

4. Conclusion and outlook

In summary, PAT for biologics has advanced in recent years towards real-time monitoring and control of critical quality

attributes. Spectroscopy based PAT tools have been successfully applied to a variety of applications. Compared to other methodologies, they feature fast measurement times, easy in-line implementation and maintainable costs.

A major challenge in future relates to a flexible implementation of PAT tools into different unit operations. Currently, disposable and single-use technologies are gaining market shares especially during clinical phases [51]. At the same time, the product portfolio of biologics is broadening. New formats such as antibody fragments, nanobodies, conjugated proteins and vaccines, and Fc-fusion proteins are emerging [52]. Depending on the unit operation and biopharmaceutical product, different sensors or sensor combinations may be of interest. Ideally, detectors could therefore be exchanged with little effort. Such a flexible approach to PAT however requires standardized communication between different components, e.g. through OPC Foundation's *OPC unified architecture* (OPC UA). Here, the support of the equipment manufacturer as well as dedicated sensor manufacturer is key [17]. By providing a flexible platform which allows to combine different manufacturing equipment with a range of sensor technologies, a versatile approach towards future PAT challenges could be implemented.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 635557.

The authors would like to thank Sebastian Andris, Dr. Pascal Bauermann and Stefan Heissler for the review of the manuscript. They are also thankful for the valuable inputs they received from other members of the research group MAB at KIT.

References

- [1] FDA, Guidance for Industry. PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, 2004.

- [2] L.L. Simon, H. Patak, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H.J.M. Kramer, J.W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K.V. Gernaey, M. Louhi-Kultanen, J. Oreilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Froberg, J. Ulrich, R.D. Braatz, T. Leyssens, M. Von Stosch, R. Oliveira, R.B.H. Tan, H. Wu, M. Khan, D. Ogrady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z.K. Nagy, J.D. Litster, V.K. Kamaraju, M.S. Chiu, Assessment of recent process analytical technology (PAT) trends: a multi-author review, *Organ. Process Res. Dev.* 19 (1) (2015) 3–62, <http://dx.doi.org/10.1021/op500261y>.
- [3] M. Schiestl, T. Stangler, C. Torella, T. Čepeljnik, H. Toll, R. Grau, Acceptable changes in quality attributes of glycosylated biopharmaceuticals, *Nat. Biotechnol.* 29 (2011) 310–312.
- [4] S. Flatman, I. Alam, J. Gerard, N. Mussa, Process analytics for purification of monoclonal antibodies, *J. Chromatogr. B* 48 (1) (2007) 79–87.
- [5] J. Glassey, K.V. Gernaey, C. Clemens, T.W. Schulz, R. Oliveira, G. Striedner, C.-F. Mandenius, Process analytical technology (PAT) for biopharmaceuticals, *Biotechnol. J.* 6 (4) (2011) 369–377.
- [6] S. Challa, R. Potumarthi, Chemometrics-based process analytical technology (PAT) tools: applications and adaptation in pharmaceutical and biopharmaceutical industries, *Appl. Biochem. Biotechnol.* 169 (1) (2013) 66–76.
- [7] S.M. Mercier, B. Diepenbroek, R.H. Wijffels, M. Streefland, Multivariate PAT solutions for biopharmaceutical cultivation: current progress and limitations, *Trends Biotechnol.* 32 (6) (2014) 329–336.
- [8] A.S. Rathore, G. Kapoor, Application of process analytical technology for downstream purification of biopharmaceuticals, *J. Chem. Technol. Biotechnol.* 90 (2) (2015) 228–236.
- [9] K.B. Konstantinov, C.L. Cooney, White paper on continuous bioprocessing. May 20–21, 2014 continuous manufacturing symposium, *J. Pharm. Sci.* 104 (3) (2015) 813–820.
- [10] N. Brestrich, T. Hahn, J. Hubbuch, Application of spectral deconvolution and inverse mechanistic modelling as a tool for root cause investigation in protein chromatography, *J. Chromatogr. A* 1437 (2016) 158–167, <http://dx.doi.org/10.1016/j.chroma.2016.02.011>.
- [11] R.L. Fahrner, P.M. Lester, G.S. Blank, D.H. Reifsnnyder, Real-time control of purified product collection during chromatography of recombinant human insulin-like growth factor-I using an on-line assay, *J. Chromatogr. A* 827 (1) (1998) 37–43.
- [12] R.L. Fahrner, G.S. Blank, Real-time control of antibody loading during protein A affinity chromatography using an on-line assay, *J. Chromatogr. A* 849 (1) (1999) 191–196.
- [13] A.S. Rathore, M. Yu, S. Yeboah, A. Sharma, Case study and application of process analytical technology (PAT) towards bioprocessing: use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography, *Biotechnol. Bioeng.* 100 (2) (2008) 306–316.
- [14] A.S. Rathore, R. Wood, A. Sharma, S. Dermawan, Case study and application of process analytical technology (PAT) towards bioprocessing. II: Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography, *Biotechnol. Bioeng.* 101 (6) (2008) 1366–1374.
- [15] O. Kaltenbrunner, Y. Lu, A. Sharma, K. Lawson, T. Tresselt, Risk-benefit evaluation of on-line high-performance liquid chromatography analysis for pooling decisions in large-scale chromatography, *J. Chromatogr. A* 1241 (2012) 37–45.
- [16] M. Krättli, F. Steinebach, M. Morbidelli, Online control of the twin-column countercurrent solvent gradient process for biochromatography, *J. Chromatogr. A* 1293 (2013) 51–59, <http://dx.doi.org/10.1016/j.chroma.2013.03.069>.
- [17] K.A. Beakev, Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries, 2nd ed., John Wiley & Sons, 2010.
- [18] G. Carta, A. Jungbauer, Protein Chromatography: Process Development and Scale-up, John Wiley & Sons, 2010.
- [19] B.R. Kowalski, Chemometrics: views and propositions, *J. Chem. Inf. Model.* 15 (4) (1975) 201–203, <http://dx.doi.org/10.1021/ci60004a002>.
- [20] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceuticals: a tutorial review, *Int. J. Pharm.* 417 (1–2) (2011) 280–290, <http://dx.doi.org/10.1016/j.ijpharm.2011.02.019>.
- [21] L. Eriksson, Multi- and Megavariate Data Analysis, Umetrics AB, 2006.
- [22] O.M. Kvalheim, Interpretation of direct latent-variable projection methods and their aims and use in the analysis of multicomponent spectroscopic and chromatographic data, *Chemom. Intell. Lab. Syst.* 4 (1) (1988) 11–25, [http://dx.doi.org/10.1016/0169-7439\(88\)80009-1](http://dx.doi.org/10.1016/0169-7439(88)80009-1).
- [23] O.M. Kvalheim, T.V. Karstang, Interpretation of latent-variable regression models, *Chemom. Intell. Lab. Syst.* 7 (1) (1989) 39–51, [http://dx.doi.org/10.1016/0169-7439\(89\)80110-8](http://dx.doi.org/10.1016/0169-7439(89)80110-8).
- [24] S. Wold, J. Trygg, The PLS method – partial least squares projections to latent structures – and its applications in industrial RPD (research, development, and production), in: *PLS in Industrial RPD for Prague 1* (June), 2004, pp. 1–44, <http://dx.doi.org/10.1109/JMEMS.2011.2159097>.
- [25] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1–3) (1987) 37–52, [http://dx.doi.org/10.1016/0169-7439\(87\)80084-9](http://dx.doi.org/10.1016/0169-7439(87)80084-9).
- [26] S. Wold, A. Ruhe, H. Wold, W. Dunn, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput.* 5 (3) (1984) 735–743, [arXiv:1308.0863v1](http://arxiv.org/abs/1308.0863v1).
- [27] H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, 1991.
- [28] S. Wold, J. Trygg, A. Berglund, H. Antti, Some recent developments in PLS modeling, *Chemom. Intell. Lab. Syst.* 58 (2) (2001) 131–150, [http://dx.doi.org/10.1016/S0169-7439\(01\)00156-3](http://dx.doi.org/10.1016/S0169-7439(01)00156-3).
- [29] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2) (2001) 109–130, [http://dx.doi.org/10.1016/S0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/S0169-7439(01)00155-1), [arXiv:S0169-7439\(01\)00155-1](http://arxiv.org/abs/S0169-7439(01)00155-1).
- [30] W. Kessler, Einführung in die multivariate Datenanalyse, 1st ed., WILEY-VCH Verlag, Weinheim, 2007.
- [31] W. Jiskoot, D. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, American Association of Pharmaceutical Scientists, 2005.
- [32] A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta* 1767 (9) (2007) 1073–1101, <http://dx.doi.org/10.1016/j.bbabi.2007.06.004>.
- [33] J.R. Lakowicz, Principles of Fluorescence Spectroscopy Principles of Fluorescence Spectroscopy, Springer Science + Business Media, LLC, 2006, <http://dx.doi.org/10.1007/978-0-387-46312-4>, [arXiv:1011.1669v3](http://arxiv.org/abs/1011.1669v3).
- [34] R. Schweitzer-Stenner, Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure: a critical review, *Vib. Spectrosc.* 42 (1) (2006) 98–117.
- [35] E. Goormaghtigh, V. Cabiaux, J.-M. Ruysschaert, Determination of soluble and membrane protein structure by fourier transform infrared spectroscopy: I. Assignment of model compounds, in: *Physicochemical Methods in the Study of Biomembranes*, vol. 23, Springer, 1994, pp. 328–357.
- [36] S.K. Hansen, B. Jamali, J. Hubbuch, Selective high throughput protein quantification based on UV absorption spectra, *Biotechnol. Bioeng.* 110 (2) (2013) 448–460.
- [37] G.E. Arteaga, Y. Horimoto, E. Li-Chan, S. Nakai, Partial least-squares regression of fourth-derivative ultraviolet absorbance spectra predicts composition of protein mixtures: application to bovine caseins, *J. Agric. Food Chem.* 42 (9) (1994) 1938–1942.
- [38] S.K. Hansen, E. Skibsted, A. Staby, J. Hubbuch, A label-free methodology for selective protein quantification by means of absorption measurements, *Biotechnol. Bioeng.* 108 (11) (2011) 2661–2669.
- [39] M.-H. Kamga, H. Woo Lee, J. Liu, S. Yoon, Rapid quantification of protein mixture in chromatographic separation using multi-wavelength UV spectra, *Biotechnol. Prog.* 29 (3) (2013) 664–671.
- [40] N. Brestrich, T. Briskot, A. Osbergerhaus, J. Hubbuch, A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression, *Biotechnol. Bioeng.* 111 (7) (2014) 1365–1373.
- [41] N. Brestrich, A. Sanden, A. Kraft, K. McCann, J. Bertolini, J. Hubbuch, 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 (2015) 1406–1416.
- [42] M. Rüdert, N. Brestrich, L. Rolinger, J. Hubbuch, Real-time monitoring and control of the load phase of a protein capture step, *Biotechnol. Bioeng.* (2016), <http://dx.doi.org/10.1002/bit.26078.pdf>.
- [43] N. Brestrich, M. Rüdert, D. Büchler, J. Hubbuch, Selective protein quantification for preparative chromatography using variable pathlength UV/vis spectroscopy, *Biotechnol. Bioeng.* (2016) (submitted for publication).
- [44] F. Capito, R. Skudas, H. Kolmar, C. Hunzinger, Mid-infrared spectroscopy-based antibody aggregate quantification in cell culture fluids, *Biotechnol. J.* 8 (8) (2013) 912–917, <http://dx.doi.org/10.1002/biot.201300164>.
- [45] F. Capito, R. Skudas, H. Kolmar, B. Stanislawski, Host cell protein quantification by Fourier transform mid infrared spectroscopy (FT-MIR), *Biotechnol. Bioeng.* 110 (1) (2013) 252–259.
- [46] F. Capito, R. Skudas, H. Kolmar, C. Hunzinger, At-line mid infrared spectroscopy for monitoring downstream processing unit operations, *Process Biochem.* 50 (6) (2015) 997–1005, <http://dx.doi.org/10.1016/j.procbio.2015.03.005>.
- [47] C. Walther, S. Mayer, A. Jungbauer, A. Dürauer, Getting ready for PAT: scale up and inline monitoring of protein refolding of Npro fusion proteins, *Process Biochem.* 49 (7) (2014) 1113–1121, <http://dx.doi.org/10.1016/j.procbio.2014.03.022>.
- [48] M. Boulet-Audet, S.G. Kazarian, B. Byrne, In-column ATR-FTIR spectroscopy to monitor affinity chromatography purification of monoclonal antibodies, *Sci. Rep.* 407 (May) (2016) 7111–7122, <http://dx.doi.org/10.1038/srep30526>.
- [49] A.S. Rathore, X. Li, W. Bartkowski, A. Sharma, Y. Lu, Case study and application of process analytical technology (PAT) towards bioprocessing: use of tryptophan fluorescence as at-line tool for making pooling decisions for process chromatography, *Biotechnol. Prog.* 25 (5) (2009) 1433–1439.
- [50] Z. Yu, J.C. Reid, Y.P. Yang, Utilizing dynamic light scattering as a process analytical technology for protein folding and aggregation monitoring in vaccine manufacturing, *J. Pharm. Sci.* 102 (12) (2013) 4284–4290, <http://dx.doi.org/10.1002/jps.23746>.
- [51] A.A. Shukla, U. Gottschalk, Single-use disposable technologies for biopharmaceutical manufacturing, *Trends Biotechnol.* 31 (3) (2013) 147–154.
- [52] K. Watson, B.N. Violand, C. Gallo, S. Sun, R. Godavarti, R.S. Wright, G.R. Bolton, K.M. Sunasara, J.L. Coffman, Addressing the challenges in downstream processing today and tomorrow, *BioPharm. Int.* 24 (2011) 8–15.

- [53] E. Papamokos, E. Weber, W. Bode, R. Huber, M.W. Empie, I. Kato, M. Laskowski Jr., Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases, *J. Mol. Biol.* 158 (3) (1982) 515–537.
- [54] A.S. Rose, A.R. Bradley, Y. Valasatava, J.M. Duarte, A. Prić, P.W. Rose, Web-based molecular graphics for large complexes, in: Proceedings of the 21st International Conference on Web3D Technology, Web3D '16, ACM, New York, NY, USA, 2016, pp. 185–186, <http://dx.doi.org/10.1145/2945292.2945324>.