Application note

# Raman-based Accurate Protein Quantification in a Matrix that Interferes with UV-Vis Measurement


## Authors

Michelle Nolasco[1], Kristina Pleitt[1], Nimesh Khadka[2]

[1]BioProduction Group, Thermo Fisher Scientific, St. Louis, Missouri USA

[2]Analytical Instrument Group, Thermo Fisher Scientific, Tewksbury, Massachusetts, USA


## Significance

Ultraviolet-Visible spectroscopy (UV-Vis) is widely used for accurate quantification of concentrations of purified proteins. In cases where the matrix strongly absorbs photons of the key wavelength (280 nm), UV-Vis leads to the inaccurate quantification of protein concentration. The issue is further complicated if the matrix is dynamic, as it adds potential error to mathematical operations applied to remove the background information. This work presents a case study of an ultrafiltration/diafiltration (UF/DF) experiment with a matrix containing tryptophan. As shown in Figure 1, both in-line UV-Vis and in-line Raman were used to monitor the process. Our findings demonstrate that Raman-based estimation of protein concentration proves to be more accurate and reliable when matrix interference is present. These results serve as a paradigm, reinforcing that Raman spectroscopy, based on specific molecular signatures, is particularly suitable for monitoring complex biological processes characterized by multi-component interactions and dynamic changes.

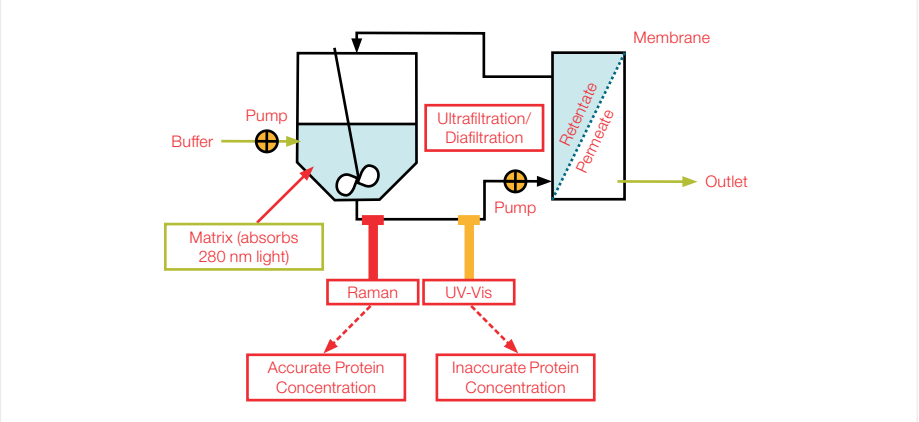



**Figure 1. Advantage of Raman based estimation of protein concentration in a UF/DF run in presence of interfering background due to matrix which varies with progress of the process.**

thermo scientific

## Introduction

The industrial biomanufacturing of therapeutic monoclonal antibodies (mAb) involves upstream and downstream processes,[1,2] as depicted in Figure 2. In the upstream process, cells are cultivated in bioreactors to produce the mAb. These mAb then undergo a series of operations in the downstream processing, including clarification, chromatography, filtration, viral inactivation, viral filtration, ultrafiltration/diafiltration (UF/DF), final filtration, and fill and finish steps. These operations are collectively categorized as downstream processing and are crucial for ensuring the purity, stability, and efficacy of the mAb.

Accurate quantification of the protein at different stages is important for downstream processes.[3,4] It enables real-time decision-making. For example, accurate quantification of the target protein after the clarification step provides actionable results such as yield estimation, GO/NO GO decisions (e.g., discontinuing a batch if the product yield falls below the threshold to save costs, time, and resources), and selection of the appropriate number of chromatography cycles required to maintain the process within expected resin loadings. Accurate protein quantification is essential during UF/DF to monitor and control the protein concentration to a predefined target. Similarly, precise protein quantification in fill and finish products ensures correct dosages are delivered to patients. Thus, accurate protein quantification is a critical process parameter for downstream processing.

The widely accepted standard method for estimating total protein concentration is at-line or in-line UV-Vis absorption spectroscopy using a light source at approximately 280 nm.[5,6] This method relies on the absorption of aromatic amino acid residues (tyrosine and tryptophan) and disulfide bonds, which act as chromophores contributing to the UV-Vis absorption.[7] In UV-Vis spectroscopy, the absorbance is linearly correlated with concentration as defined by the Beer-Lambert Law:

In cases where the matrix absorbs photons at 280 nm, the total absorbance (matrix + mAb) will be higher than the absorbance of the pure mAb for a given concentration, resulting in overprediction. If the absorbance of the matrix is low and constant, a simple correction factor can be calculated and subtracted from the total absorbance to obtain an accurate mAb concentration. However, estimating the correction factor can be challenging when the matrix is dominant and dynamic.

One example of complex matrix interference in downstream processing is the clarified harvest. The clarified harvest pool contains interfering components of various concentrations from the upstream process, such as host cell proteins, amino acids, and other biomolecules. This prevents the use of UV-Vis to estimate mAb concentration. Another example of matrix interference is diafiltration buffer excipients that absorb UV-Vis photons, leading to overestimating protein concentration during UF/DF. In both these examples, mAb is first purified using affinity chromatography (Protein A). Then, the concentration is determined by reading through the at-line UV-Vis instrument (i.e., an HPLC titer assay). This additional sample preparation and analytical steps increase the cost and process time from a couple of hours to potentially days, depending on the workload.

An alternative technology that can be used instead of UV-Vis for downstream monitoring, particularly in situations where there is a matrix effect, is process Raman. Raman spectroscopy is an optical technique that involves measuring the vibrational modes of molecules by observing the inelastic scattering of photons (known as Raman scattering) after interacting with a light wave.[8] It is highly specific for the target molecules and, thus, widely used for identification and quantification. In this study, we present a case study where we used process Raman and in-line UV-Vis to monitor a UF/DF run with a UV-Vis interfering matrix and demonstrate the advantage of Raman for such applications.

$$Concentration = \frac{Absorbance}{path\ length * molar\ extinction\ coefficient}$$

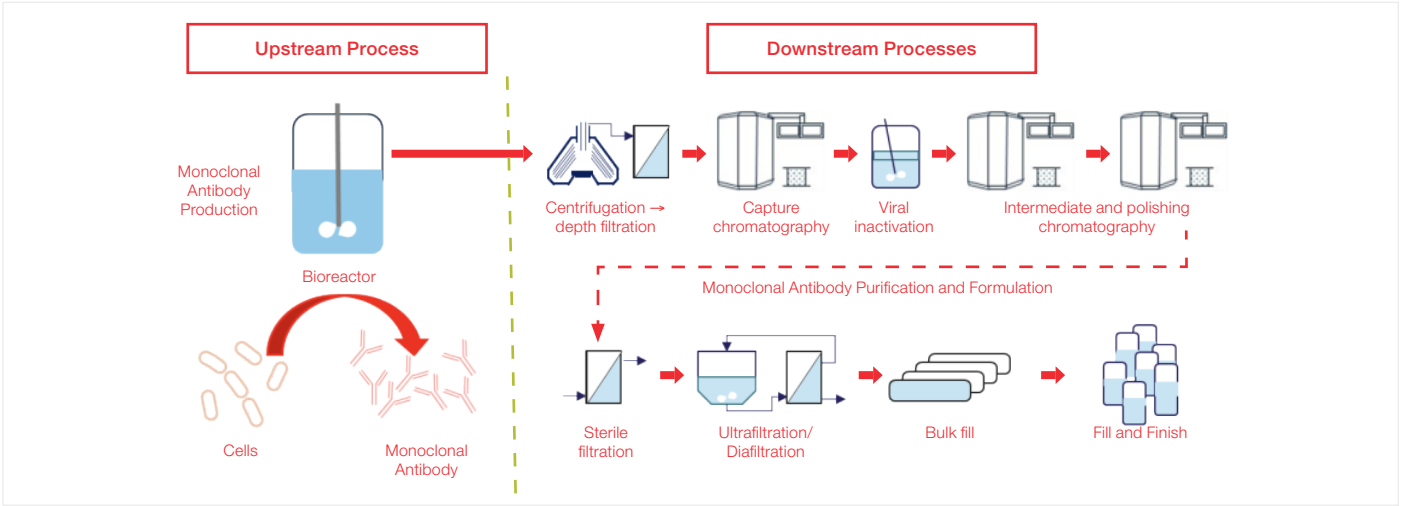
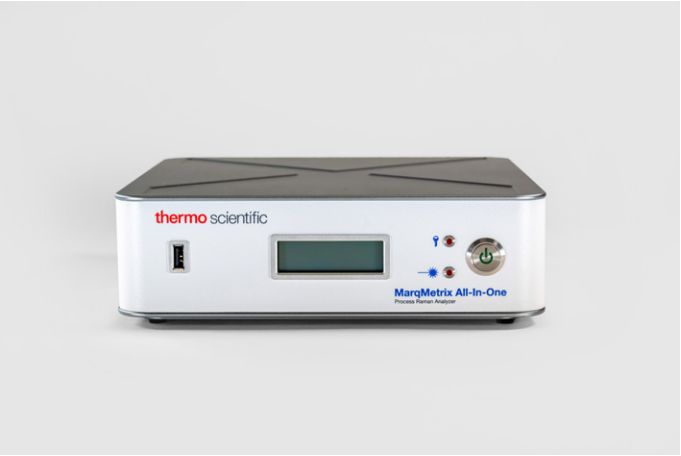



**Figure 2. Upstream and downstream processes for monoclonal antibody production.**




## Experimental details

### Calibration model development

The protein quantification partial least square (PLS) models for Raman were developed using the spectra collected on a single monoclonal antibody (concentration range 0 to 157 mg/mL measured at-line using UV-Vis instrument) in various buffer and excipient-containing solutions. The samples were passed through the flowcell probe integrated with Thermo Scientific™ MarqMetrix™ All-In-One Process Raman Analyzer at a 100 mL/min flow rate. The Raman spectra were acquired using a 785 nm laser with the following acquisition parameters: laser power 450 mW, integration time 3,000 ms, an average of 3 (i.e., a single spectrum per 18 s), and ten replicates per concentration.

As preprocessing steps, each spectrum was normalized by dividing it by the infinity norm, which was calculated using the water band of each spectrum using region 3,098 to 3,230 $cm^{-1}$. The normalized spectra were further processed with the SavGol filter (1st derivative, order = 2, window width = 13) and mean centering. The PLS model was developed using the spectral region 1,600 to 1,750 $cm^{-1}$ and water band region 3,098 to 3,230 $cm^{-1}$. The Raman spectral region 1,600 to 1,750 $cm^{-1}$ of mAb is assigned to the vibration of the carbonyl group in the amide bond (-CO-NH-) located at different secondary structures like α-helix, β-sheet, turns, and random coils.[9] Thus, the model is termed the “Amide I model.” Another PLS model was developed using the same preprocessing but utilizing the spectral region 800 to 1,750 $cm^{-1}$ and water band region 3,098 to 3,230 $cm^{-1}$, which we termed the “Extended Region model.” The models were internally validated using a leave-one-out cross validation strategy. The variable importance in projection (VIP) scores was calculated for each model to calculate the importance of each Raman shift for the development of the model.[10] It also allows us to validate the statistical model with chemical information. The details on strategies for model development, validation, and VIP analysis were discussed previously. All chemometric works were performed using SOLO 9.3.1 (2024). Eigenvector Research. Inc. Manson, WA USA 98831.

### Tryptophan as an excipient in the ultrafiltration/ diafiltration (UF/DF)

Ultrafiltration diafiltration (UF/DF) is a common unit operation in downstream bioprocessing involving a known pore-size tangential flow filtration membrane. It is used to buffer exchange and concentrate the desired biomolecules to prepare the product for final formulation. In this study, we started the experiment with ~10 mg/mL mAb in tris buffer and concentrated it to ~23 mg/mL in tris buffer. At this point, we initiated diafiltration to exchange the tris buffer with the excipient buffer. One of the components of the excipient buffer was tryptophan (20mM). The mAb concentration throughout the UF/DF run was monitored using an in-line process Raman and in-line UV-Vis spectrometer, as shown in Figure 1. The reference mAb concentration at the different stages of the UF/DF run was quantified by analytical protein A titer assay. The in-line Raman predicted mAb concentration and at-line HPLC-titer measured mAb concentration were used to calculate the root mean square error of prediction (RMSEP).

### Test of model transferability

The Amide I and Extended Region calibration models were developed without incorporating any training dataset on tryptophan. Instead, the training dataset for both models consisted of mAb (IgG 4) samples in various buffers, including tris, histidine, arginine, sucrose, and polysorbate. As a result, the predictive performance of these models also assesses their transferability across different processes.

## Results

The results of the PLS calibration models for predicting protein concentration based on Raman data are shown in Figure 3. The details of the strategy of model development, validation, and the VIP plot have been discussed in previous work. Two latent variables were selected to build the Amide I and Extended Region PLS model. The root means square error of cross-validation (RMSECV) of approximately 0.70 mg/mL for both models (Figure 3A and 3C inset) across the training concentration of 0 to 157 mg/mL mAb indicates that both models have high accuracy. The $R^2$ for cross-validation being close to 1 for both models indicates that the models explain the variance in the training data related to mAb very well. The VIP score plot shows the importance of the Raman shift for the model. Here, the VIP score one is defined as the threshold (red dotted line in Figure 3B and 3D). Any Raman shifts with a VIP score of more than one are considered important for the model.

Similarly, the higher the scores, the more important the Raman shift is for the model. For the Amide I model, the Raman shift at ~1,670 $cm^{-1}$ is influential in the model (Figure 3B). On the other hand, the amide region (~1,670 $cm^{-1}$), CH deformation (~1,450 $cm^{-1}$), breathing mode of phenylalanine (~1,005 $cm^{-1}$), and fermi doublet of tyrosine (~830-850 $cm^{-1}$) are dominant in the Extended Region model (Figure 3D).

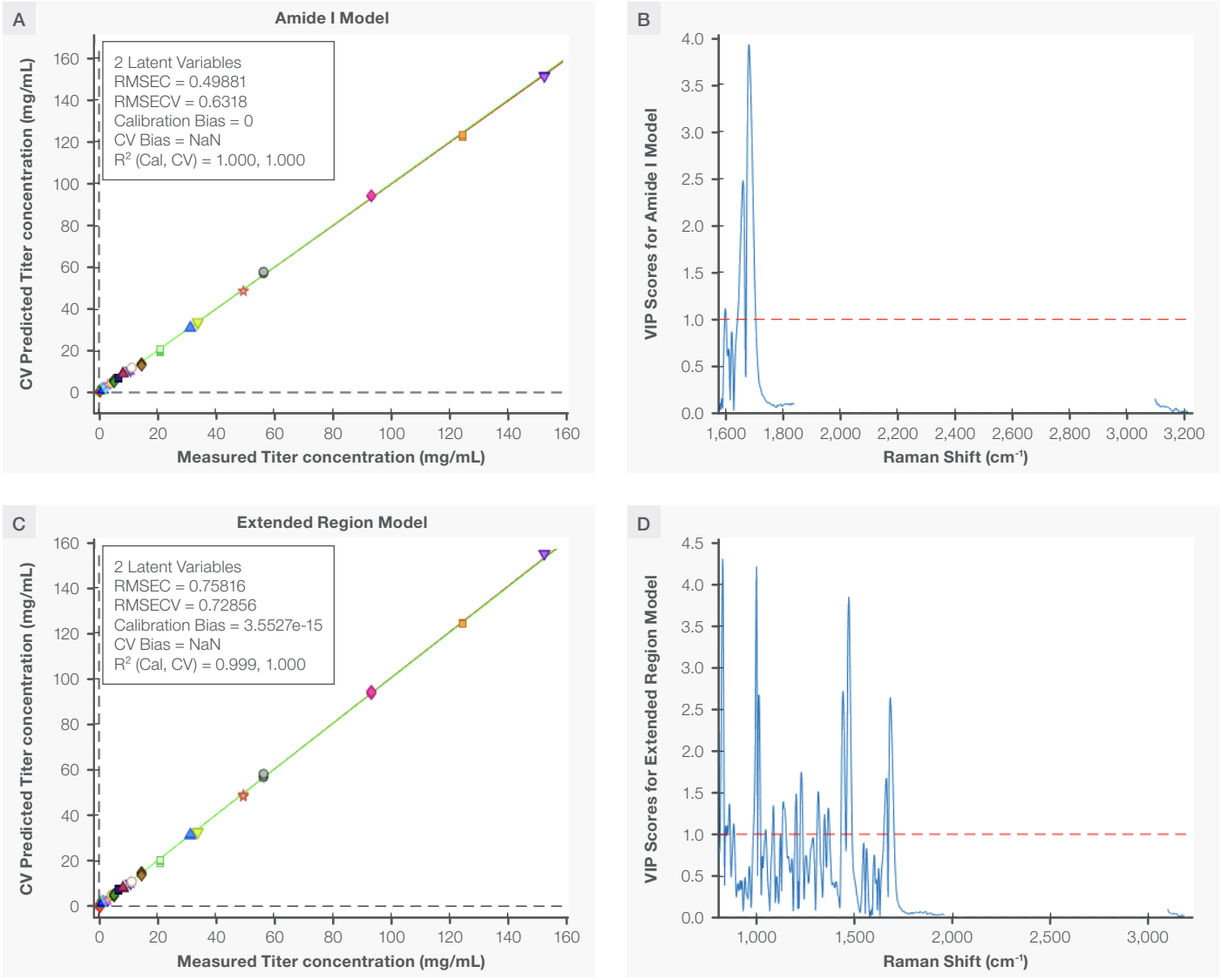



**Figure 3. Plots A and B show the Amide I calibration model and its VIP score plot, respectively. Similarly, plots C and D show the Extended Region calibration model and its VIP score plot. The model statistics are shown in the inset of plots A and C.**

The results for the UF/DF run are shown in Figure 4. The process was initiated with 10 mg/mL mAb in tris buffer and concentrated to 23 mg/mL in tris buffer. In this concentration step, the prediction of mAb concentration from in-line UV-Vis and in-line Raman (for both models) showed excellent agreement with an overall absolute prediction error of less than 5%. In the diafiltration step, the mAb in the tris buffer was exchanged with the buffer containing tryptophan. As the tryptophan was introduced into the system, the prediction of mAb concentration from in-line UV-Vis had an absolute prediction error of 20 to 70% compared to the HPLC-titer (Figure 4). In contrast, the real-time predictions from in-line process Raman using Amide I and Extended Region model were more accurate. The absolute prediction error for the Amide I model was below 5%, while for the Extended Region model, it was below 10% (Figure 5).

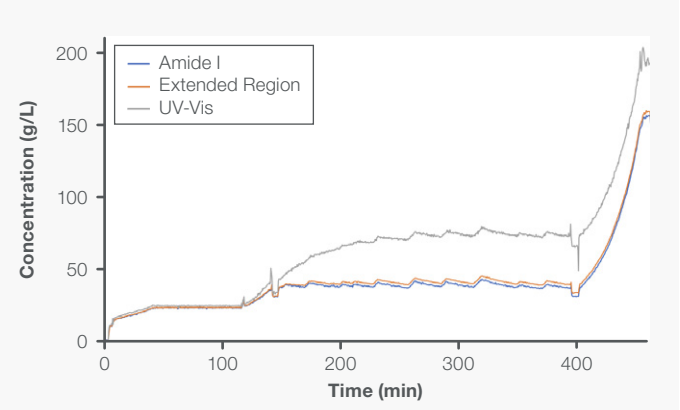



**Figure 4. Overlay of mAb predictions from in-line UV-Vis (grey) and in-line process Raman (blue and orange) in the UF/DF run. The blue and orange traces represents the Raman predictions of the mAb concentration using Amide I and Extended Region models respectively. The excipient buffer containing tryptophan was introduced in the system at time ~120 min. The mAb prediction using in-line UV-Vis and in-line Raman were accurate and similar (<5% differences) before tryptophan was introduced (0 to 120 min). After introducing tryptophan, the predicted error form in-line UV-Vis was between 20 to 70%.**

## Why are Raman predictions accurate?

The accuracy of Raman predictions can be explained based on molecular specificity, which forms the fundamental basis of all Raman applications. Raman is based on the unique signature exhibited by molecules with different atomic compositions or molecular bonds. In this study, the molecular specificity for mAb is provided by the Raman signature of the Amide I region that extends from ~1,640 – 1,700 $cm^{-1}$. Amide bonds (-CO-NH-) are absent in this study's tryptophan and other buffer components. This is evident in the second derivative plot of mAb concentration (~ 31 mg/mL) in tris buffer (red) and excipient buffer with tryptophan (green), as shown in Figure 6A. The spectral overlay of the Amide I region with and without tryptophan is approximately identical in intensity (at ~1,670 $cm^{-1}$), and peak features indicate no spectral interference. As shown in the VIP plot for the Amide I model (Figure 3B), the 1,670 $cm^{-1}$ region is dominant in the calibration model and crucial in making predictions. This unique Raman signature of the Amide I region of mAb explains the accurate transferability of the Amide I model with an absolute prediction error of less than 5% (Figure 5).

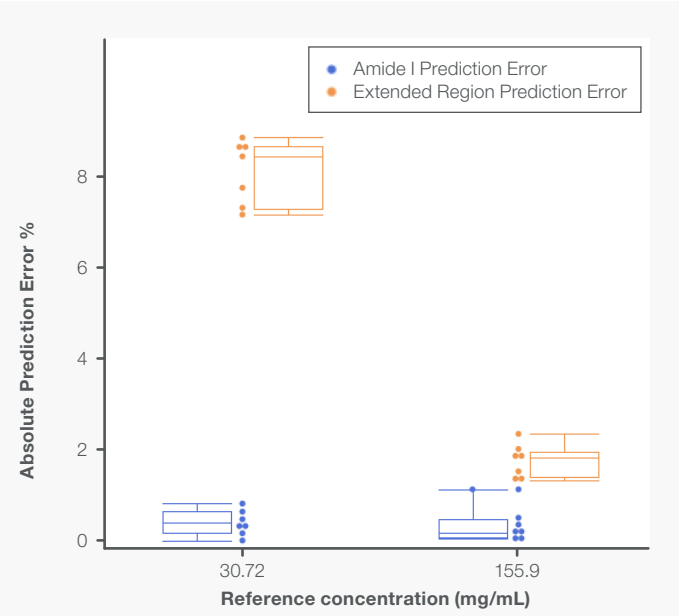



**Figure 5. Plot showing absolute error for mAb predictions from Amide I and Extended Region models at the reference concentration of mAb of 30.72 and 155.9 mg/mL in presence of 20mM tryptophan. *Note, none of the tryptophan dataset were included in the training set for either model.***

Tryptophan, an amino acid part of mAb, has a strong Raman signal that overlaps the protein spectra beyond the Amide I region, as shown in Figure 6B. The Raman peaks from tryptophan at ~1,623 $cm^{-1}$ and ~1,555 $cm^{-1}$ are assigned to the stretching vibration of the benzene ring; the peak at ~1,436 $cm^{-1}$ to the stretching vibration of the pyrrole ring; and at ~882 $cm^{-1}$ to the skeletal vibration, with significant contribution from the pyrrole NH in-plane deformation.[11] However, the prediction errors from the Extended Region model were less than 10% (Figure 5) due to the molecular specificity provided by including the Amide I region in the calibration model. The slightly higher error in the prediction from the Extended Region Model compared to the Amide I Model is mainly due to spectra overlap with tryptophan. By augmenting the Extended Region model with an appropriate tryptophan dataset using the design of the experiment approach, the prediction error could be improved further.

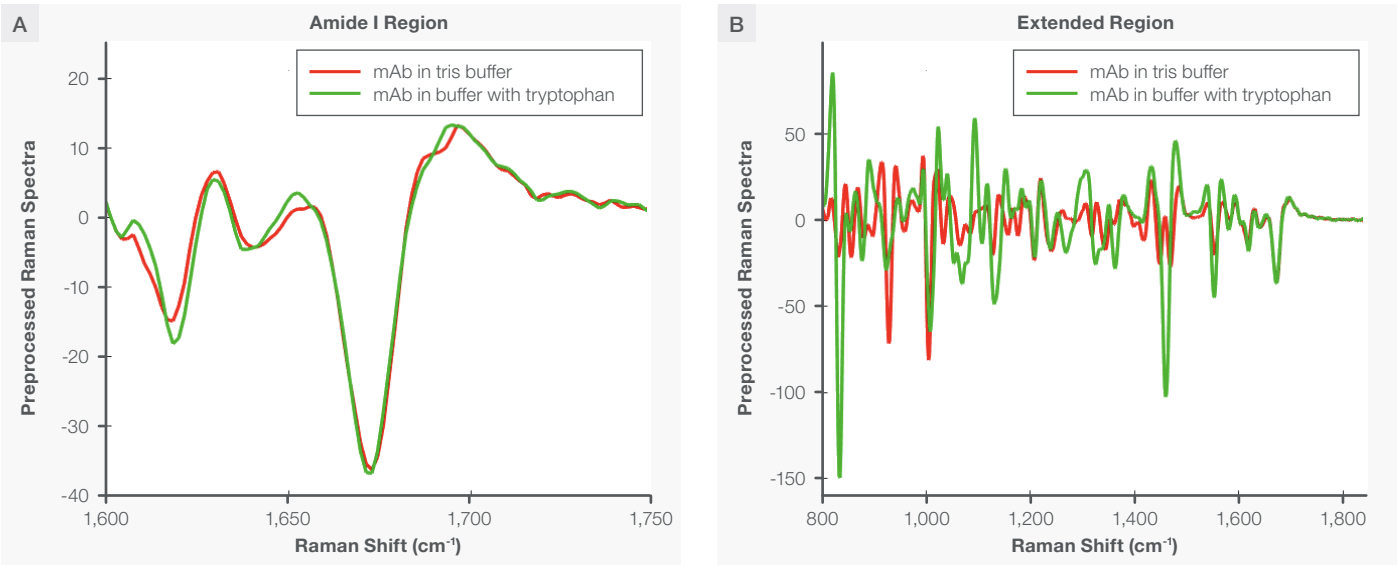



**Figure 6. Spectral overlap of mAb in tris (red) and tryptophan containing buffer (green) after applying SavGol filter (2nd derivative). The Amide I region is free of spectral overlap, thus providing specificity to the models.**

## Conclusion

- We presented a case study demonstrating the advantage of in-line process Raman for accurately estimating mAb concentration in downstream processes, specifically UF/DF, where tryptophan interfered with direct in-line UV-Vis analysis. Raman spectroscopy overcame this interference and provided reliable and real-time mAb concentration predictions.
- Our study showcased the transferability of Raman models across processes with high accuracy. This transferability is attributed to the fact that Raman spectroscopy relies on specific molecular signatures, making it highly specific for the target analytes. This characteristic of Raman spectroscopy makes it particularly suitable for monitoring and controlling complex and dynamically changing biological processes.
- In downstream processing for mAb production, the analytes are typically present at high concentrations and purity levels. This makes Raman spectroscopy an ideal tool for monitoring and controlling these processes. Not only can Raman quantify different analytes accurately, it also provides valuable insights into critical quality attributes. The unique advantage of process Raman is its ability to provide real-time and actionable results by measuring multiple analytes in a single scan. This capability aligns with the objectives outlined in the FDA guidance for Process Analytical Technology (PAT). It accelerates progress towards reducing batch-to-batch variability and ensuring uniformity in the quality of products.

## References


1. Bergemann, K.; Eckermann, C.; Garidel, P.; Grammatikos, S.; Jacobi, A.; Kaufmann, H.; Kempken, R.; Pisch-Heberle, S. Production and Downstream Processing. In *Handbook of Therapeutic Antibodies*; John Wiley & Sons, Ltd, **2007**; pp 199–237. https://doi.org/10.1002/9783527619740.ch9.
2. Gronemeyer, P.; Ditz, R.; Strube, J. Trends in Upstream and Downstream Process Development for Antibody Manufacturing. *Bioengineering* **2014**, *1* (4), 188–212. https://doi.org/10.3390/bioengineering1040188.
3. Rolinger, L.; Rüdt, M.; Diehm, J.; Chow-Hubbertz, J.; Heitmann, M.; Schleper, S.; Hubbuch, J. Multi-Attribute PAT for UF/DF of Proteins—Monitoring Concentration, Particle Sizes, and Buffer Exchange. *Anal. Bioanal. Chem.* **2020**, *412* (9), 2123–2136. https://doi.org/10.1007/s00216-019-02318-8.
4. Rathore, A. S.; Bhambure, R.; Ghare, V. Process Analytical Technology (PAT) for Biopharmaceutical Products. *Anal. Bioanal. Chem.* **2010**, *398* (1), 137–154. https://doi.org/10.1007/s00216-010-3781-x.
5. Brestrich, N.; Rüdt, M.; Büchler, D.; Hubbuch, J. Selective Protein Quantification for Preparative Chromatography Using Variable Pathlength UV/Vis Spectroscopy and Partial Least Squares Regression. *Chem. Eng. Sci.* **2018**, *176*, 157–164. https://doi.org/10.1016/j.ces.2017.10.030.
6. McKechnie, W. S.; Tugcu, N.; Kandula, S. Accurate and Rapid Protein Concentration Measurement of In-Process, High Concentration Protein Pools. *Biotechnol. Prog.* **2018**, *34* (5), 1234–1241. https://doi.org/10.1002/btpr.2695.
7. Biter, A. B.; Pollet, J.; Chen, W.-H.; Strych, U.; Hotez, P. J.; Bottazzi, M. E. A Method to Probe Protein Structure from UV Absorbance Spectra. *Anal. Biochem.* **2019**, *587*, 113450. https://doi.org/10.1016/j.ab.2019.113450.
8. Mulvaney, S. P.; Keating, C. D. Raman Spectroscopy. *Anal. Chem.* **2000**, *72* (12), 145–158. https://doi.org/10.1021/a10000155.
9. Dolui, S.; Mondal, A.; Roy, A.; Pal, U.; Das, S.; Saha, A.; Maiti, N. C. Order, Disorder, and Reorder State of Lysozyme: Aggregation Mechanism by Raman Spectroscopy. *J. Phys. Chem. B* **2020**, *124* (1), 50–60. https://doi.org/10.1021/acs.jpcb.9b09139.
10. Chong, I.-G.; Jun, C.-H. Performance of Some Variable Selection Methods When Multicollinearity Is Present. *Chemom. Intell. Lab. Syst.* **2005**, *78* (1), 103–112. https://doi.org/10.1016/j.chemolab.2004.12.011.
11. *Characterization of a few Raman lines of tryptophan - Hirakawa - 1978 - Journal of Raman Spectroscopy - Wiley Online Library.* https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/jrs.1250070511 (accessed 2024-06-30).


Learn more at **thermofisher.com/marqmetrixAIO**