A single concentration-based amide I model for real-time protein quantification and quality assessment in ultrafiltration/diafiltration using Raman spectroscopy

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

[1] Analytical Instrument, Thermo Fisher Scientific, Tewksbury, MA, USA | Thermo Fisher Scientific, St. Louis, MO, USA

Abstract

We present a Raman spectroscopy–based Classical Least Squares (CLS) chemometric approach for rapid protein quantification during downstream ultrafiltration/diafiltration (UF/DF) processing. By selecting a single reference spectrum and focusing on the protein-specific amide I Raman region normalized to the water stretching O–H band, a CLS model was developed using minimal training data to translate Raman intensities into protein concentration predictions. The model demonstrated excellent performance over a wide concentration range (0–155 mg/mL) with a root mean square error of prediction (RMSEP) of approximately 2.25 mg/mL and strong linearity (R^2 ~ 0.997), showing scalability from laboratory to pilot scale and transferability across different monoclonal antibodies and buffer matrices. In addition to concentration monitoring, the amide I region provides molecular insight into protein secondary structure, enabling simultaneous assessment of protein quantity and quality. This single-spectrum CLS strategy reduces calibration burden with cost and time benefits while delivering accuracy within typical PAT tolerances (≤5–10%), making Raman spectroscopy a practical and efficient solution for in-line process monitoring and control in biopharmaceutical manufacturing.

Motivation

- UF/DF requires accurate concentration tracking to reach targets without over-processing.
- Raman spectroscopy is a strong PAT candidate, but traditional multivariate models usually require large calibration sets.
- This work demonstrates a chemically informed CLS strategy that can be initialized from a single known-concentration spectrum, enabling rapid deployment.

Objectives

Rapid, interpretable in-line Raman quantification of mAb concentration.
Objectives:

1. Single concentration CLS model development
2. Validate across concentration and matrices
3. Demonstrate transfer to UF/DF pilot run
4. Provide performance w.r.t PLS model

Methodology

A monoclonal antibody (93.56 mg/mL) in a histidine, arginine, and sucrose buffer was flowed through a Thermo Scientific™ MarqMetrix™ FlowCell™ Sampling Optic at 100 mL/min. Raman spectra were collected in-line using a Thermo Scientific™ MarqMetrix™ All-In-One Process Raman Analyzer (450 mW, 3000 ms, 3 averages); ten spectra were denoised and averaged. Two Raman regions (3100–3230 cm^{-1}, water; 1570–1750 cm^{-1}, amide I) were used to build a CLS model. Spectra were normalized using the water band infinity norm and baseline-corrected with a Savitzky–Golay second derivative. The resulting ratio-metric CLS model predicts protein concentration from the amide I to water intensity ratio. Data processing was done in Python™, and chemometric modeling in SOLO 9.3.1 (2024) Eigenvector Research. Inc., Manson, WA USA 98831.

Figure 1. PAT integration schematic (UF/DF + Raman).

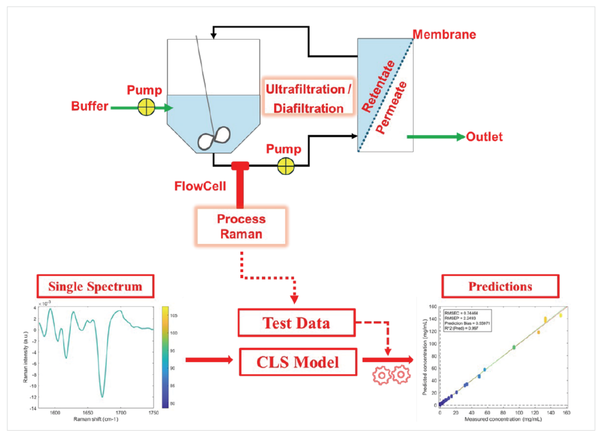


Figure 2. CLS model workflow.

1. **Model setup**
   y = Xb + e where,
   X represents the preprocessed Raman spectra
   y represents calibration concentration (93.56 mg/mL).
   b is the CLS regression vector estimated during calibration.
   e represents the residual error.

2. **Calibration step**
   $\hat{b} = (X^T X)^{+} X^T y$ where,
   regression vector b has units of mg/mL per unit
   Raman intensity.

3. **Prediction for a new spectrum**
   $\hat{y}_{new} = x_{new}^T \hat{b}$ where,
   x_new is the preprocessed Raman spectrum of the
   unknown sample.

Figure 3. CLS model development.

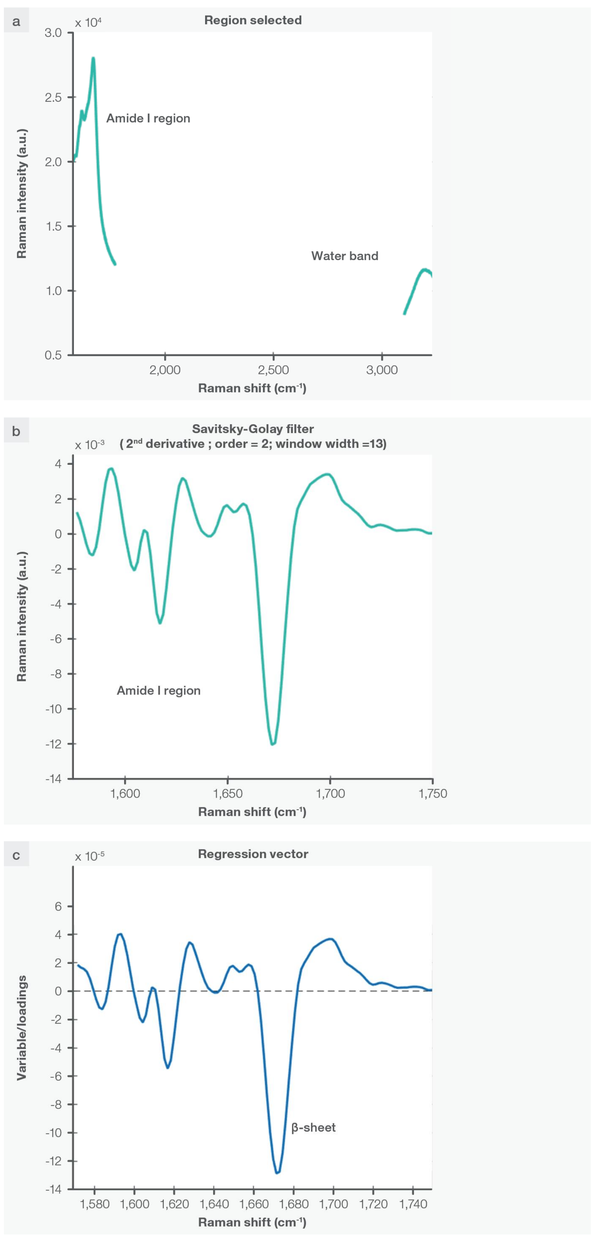


Figure 4. CLS model performance on lab scale.

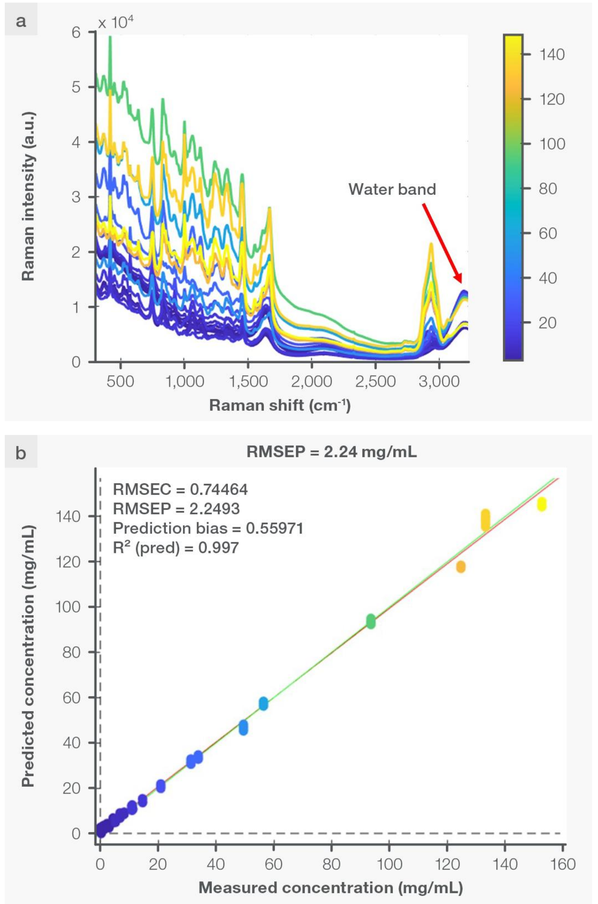
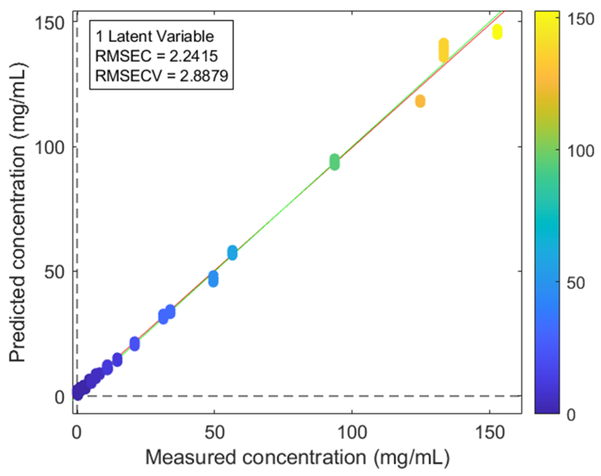
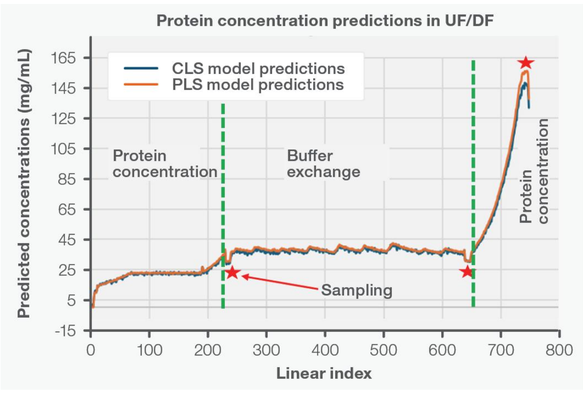


Figure 5. PLS model development



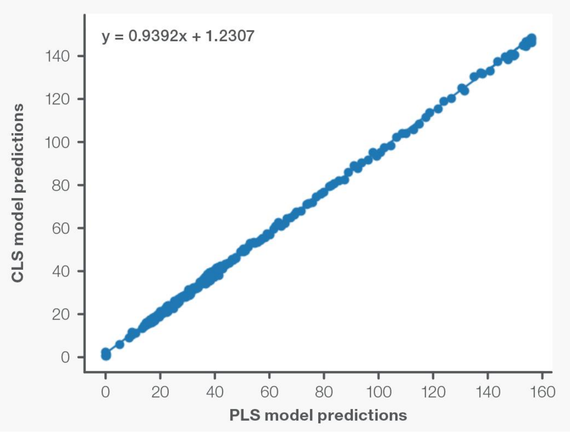
Spectral region shown in Figure 3a was selected for the data shown in figure 4a used to develop PLS model. Normalization was performed as done for the CLS model then passed through Savitzky-Golay filter (1st derivative; order =2 ; window = 13) followed by mean centering. LOOCV result is shown.

Figure 6. CLS model performance on UF/DF pilot scale with different mAb in histidine, arginine, sucrose and polysorbate matrix.



Reference (mg/mL)	PLS model predictions (mg/mL)	CLS model prediction (mg/mL)
30.72	30.65 ± 0.4	30.75 ± 0.1
155.9	154.86 ± 1.0	148.37 ± 1.4

Figure 7. CLS vs PLS performance



Summary statistics	
Mean difference	-1.18
Std dev of difference	1.69
Sample size (n)	747
Std error	0.06
T statistic	-19.08
Degrees of freedom	756
Two-tailed p-value	0.00

Conclusions

A single concentration CLS model-based Raman approach was demonstrated to enable protein quantification in downstream processing. The prediction errors were well below the < 5–10 % tolerance for PAT monitoring and showed scalability from laboratory to pilot scale as well as transferability across monoclonal antibodies and buffer matrices. Chemically informed spectral region selection was critical to model performance; the amide I Raman band provided high specificity for monoclonal antibodies with minimal spectral interference, enabling a robust and interpretable CLS model that scaled linearly with protein concentration.

References

Khadka N, Pleitt K, Nolasco M. A Classical Least Squares (CLS) Approach for Protein Quantification in Downstream Processing Using Raman Spectroscopy. Thermo Fisher Scientific Technical Note TN1384 (2025).

Trademarks/licensing

Science at a scan
Scan the QR code on the right with your mobile device to download this and many more scientific posters.

Learn more at **thermofisher.com** or contact Nimesh Khadka at **nimesh.khadka@thermofisher.com**