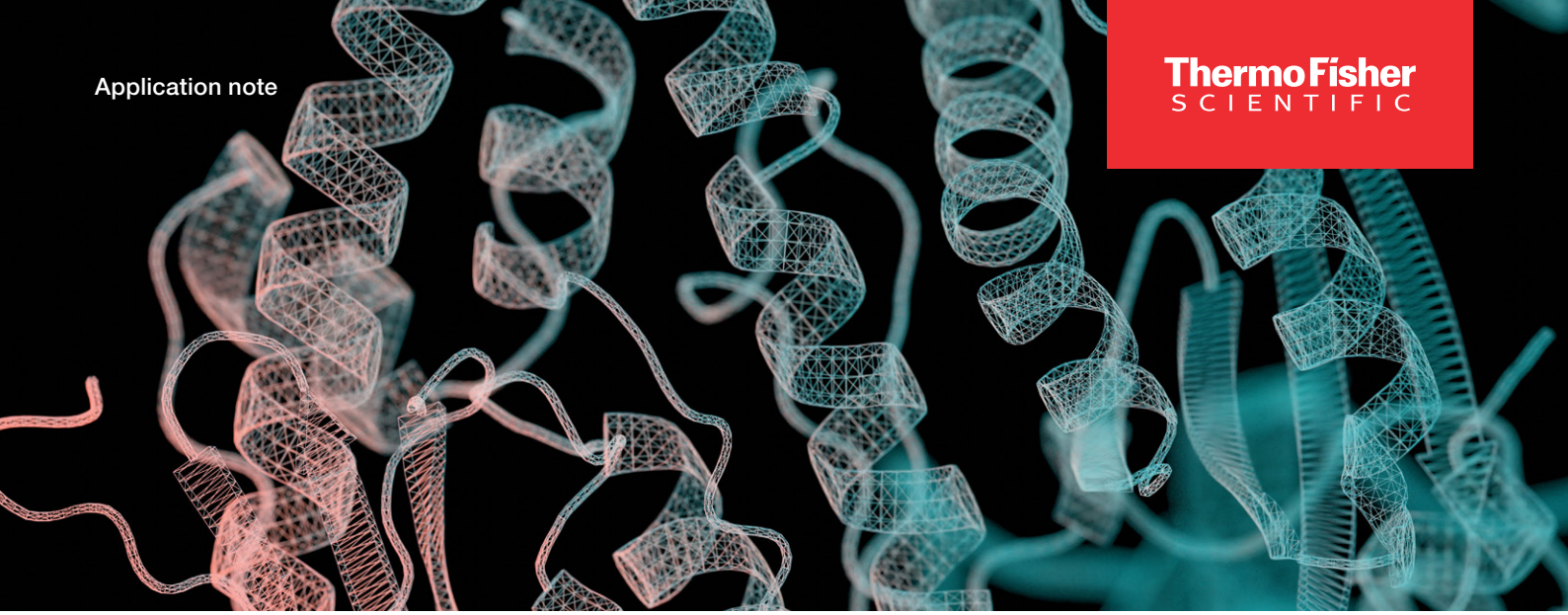


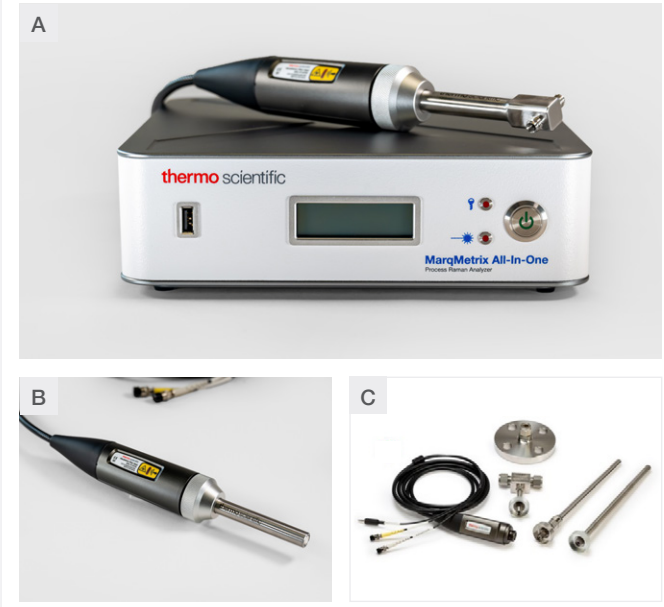
Application note

Protein secondary structure analysis using in-line process Raman and offline SERS surface

Authors

Nimesh Khadka, Ph.D.
Bin Han, Ph.D.

Analytical Instrument Group, ThermoFisher Scientific, Tewksbury, Massachusetts USA



A. Thermo Scientific MarqMetrix All-In-One Process Raman Analyzer with Thermo Scientific MarqMetrix FlowCell Sampling Optic.

B. Thermo Scientific MarqMetrixBallProbe Sampling Optic

C. Easy swapable fiber head and other sampling probes

Industry/Application:

Biopharma PAT / R&D laboratories / Downstream

Products used:

Thermo Scientific™ MarqMetrix™ All-In-One Process Raman Analyzer, Thermo Scientific™ MarqMetrix™ FlowCell Sampling Optic, Thermo Scientific™ MarqMetrix™ BallProbe™ Sampling Optic, Thermo Scientific™ OMNIC™ Software Suite

Goals:

- Demonstrate the capability of the MarqMetrix All-In-One Process Raman analyzer to acquire high-quality data for in-line protein secondary structure determination.
- Highlight the easy probe swap feature, facilitated by an optical fiber cable, which allows for seamless deployment for offline uses. In this case, SERS measurements to determine protein structure in samples in low concentration samples are featured.

Key Analytes:

Lysozyme, protein secondary structure

Key Benefits:

- Process Raman provides real-time insight into protein secondary structure during downstream purification that ensures quality of products.
- Process Raman's ability to provide simultaneous information on protein concentration and its quality makes it a valuable process analytical technology (PAT) tool.
- Use of process Raman for offline SERS measurements enable users to leverage MarqMetrix All-In-One Process Raman analyzer for all mode of applications: in-line, at-line, online, or offline.

thermo scientific

Introduction

Proteins are made up of amino acids linked together by peptide bonds (amide bonds; -CO-(NH)-), which are formed when the α-carboxyl group of one amino acid reacts with the α-amino group of another, releasing a water molecule in the process.[1] This sequence of amino acids is called the primary structure. The primary structure can fold spontaneously, or with the help of molecular chaperones, into various secondary structures such as α-helices, β-sheets, β-turns, and 3_{10} helices. These secondary structures then interact in three-dimensional space to form the protein's tertiary structure. For proteins with multiple subunits, these individual units come together to form a multimeric three-dimensional complex through various interactions known as the quaternary structure.

Protein secondary structures are crucial for determining the functions of proteins because they maintain the structural integrity of the molecules. Research has shown that tracking changes in these structures can reveal functional losses due to conformational changes, degradation, denaturation, and aggregation of proteins. Although X-ray crystallography and cryo-electron microscopy (cryo-EM) provide accurate information on protein secondary structures, they are impractical for routine use in protein biomanufacturing due to their high costs, lengthy processing times, and resource demands.[2] A more practical and efficient solution is optical analysis. Techniques like Fourier transform infrared spectroscopy (FTIR) and Raman spectroscopy have proven effective for quickly analyzing protein secondary structures in offline operation mode.[3] For routine assessments, the spectroscopic data are processed using multivariate analyses to create chemometric models. These models are validated by comparing their results with established secondary structure information from X-ray crystallography or cryo-EM. Once confirmed for accuracy, the models can be used for rapid analysis of future batches of protein samples, enabling users to evaluate protein secondary structure in a laboratory setting.[4]

This study presents the feasibility of using the in-line Thermo Scientific™ MarqMetrix™ All-In-One Process Raman Analyzer for real-time protein secondary structure analysis during downstream processing. Our previous works have already demonstrated accurate protein quantification in clarified harvest and ultrafiltration/diafiltration (UF/DF), as well as excipients quantification in downstream workflow.[5–7] Thus, this work complements our previous findings and establishes process Raman as a reliable PAT tool for multicomponent monitoring and multi-modal feedback to enable tighter process control. Additionally, this report discusses the integration of process Raman analyzer with an internally developed Surface-Enhanced Raman Spectroscopy (SERS) solid substrate as an offline alternative for determining secondary structures in samples with low protein concentrations.

Experimental Details

Data collection for initial proof of concept (PoC)

A 17 mg/mL lysozyme solution was prepared in water. Approximately 1 mL of this solution was flushed through the Thermo Scientific™ MarqMetrix™ FlowCell Sampling Optic, which has a sampling volume of approximately 180 µL. The residual lysozyme solution inside the FlowCell after flushing was used to collect the Raman spectrum. The FlowCell was connected to the MarqMetrix All-In-One Process Raman analyzer via an optical fiber. Raman data were acquired using a laser power of 450 mW, an integration time of 5000 ms, and an average of 10 scans. Three acquired spectra were further averaged to improve the signal-to-noise ratio (SNR). The same strategy and acquisition parameters were used to acquire Raman spectra of water.

Cation exchange purification

A buffered solution of 3.95 mg/mL lysozyme in 20mM MES buffer pH 6.2 was loaded at 0.83 ml/mL into a column packed with POROS™ 50 HS Strong Cation Exchange Resin. After loading, the column was washed with 5 column-volumes of 20mM MES buffer pH 6.2. Finally, the bound lysozyme was eluted at a flow rate of 3mL/min using elution buffer (20mM MES buffer 1M NaCl pH 6.2). This entire chromatographic run was monitored with the UV-Vis detector of the AKTA Pure (300 nm wavelength) and MarqMetrix All-In-One Process Raman analyzer integrated with a FlowCell probe. The Raman spectra were acquired using the acquisition settings of laser power 450 mW, integration time 3000 ms, and average of 3. During the elution step, the lysozymes are concentrated and eluted from the column. When the concentrated lysozyme reached the FlowCell cavity, the flow was paused and the Raman spectrum was acquired with the acquisition setting of power 450 mW, integration time 5000 ms, and average of 10. Three of these spectra were averaged for further analysis.

SERS data collection

A 5 µL aliquot of 0.1 mg/mL lysozyme in water was added on the SERS surface. The sample was allowed to concentrate through evaporation for approximately 10 min. This step provides enhancement due to both the increased concentration and true SERS effect. Raman spectra were collected using a MarqMetrix All-In-One Process Raman analyzer integrated with the Thermo Scientific™ MarqMetrix™ Proximal BallProbe™ Sampling Optic via optical fiber. The Raman data were acquired using a laser power of 200 mW, an integration time of 5000 ms, and an average of 10 scans after optimizing the focal distance using XYZ three axis micrometer stage to obtain the maximum signal.

Data analysis and peak deconvolution

For the initial proof-of-concept (PoC) study, the Raman spectra of water and a 17 mg/mL lysozyme solution in water were smoothed using a Savitzky-Golay filter (window width = 7, order = 2) and then each spectrum was then normalized using the weight vector that was obtained by calculating the infinity norm in the region of 3000 to 3240 cm^{-1}, which corresponds to the symmetric stretching of O-H bonds in water molecules. This region was chosen for normalization as it has minimal spectral interferences from the Raman signatures of lysozyme or buffer analytes, and thus can correct for any path length differences.

After normalization, the spectral region from 1400 to 1800 cm^{-1} was selected for each spectrum. The baseline was removed from each spectrum using an Automatic Whittaker filter (lambda = 5000, asymmetry (p) = 0.001). The baseline-removed water Raman spectrum was then subtracted from the baseline-removed lysozyme solution Raman spectrum to isolate the pure lysozyme spectrum.

This pure lysozyme Raman spectrum was preprocessed using a Savitzky-Golay filter (window width = 9, order = 2, second derivative) to identify the number and positions of peaks. After identifying the peaks, the pure lysozyme Raman spectrum was exported as a .spc file. All data analysis explained above were performed using the SOLO 9.3.1 software package (2024, Eigenvector Research, Inc., Manson, WA, USA 98831).

The exported .spc file was loaded into Thermo Scientific™ OMNIC™ Software for peak deconvolution in the spectral region of approximately 1500 to 1800 cm^{-1}. The Voigt function was selected for peak fitting. The initial guesses for the number of peaks and their positions were defined based on previously calculated second derivative data. The initial guess for the full width at half maximum (FWHM) for the Voigt peaks was set at 8 cm^{-1}, considering the resolution of instrument was about 6 cm^{-1}. The noise level was calculated using the spectral range of 1750 to 1780 cm^{-1}. With these initial guesses, the global optimization was performed in the OMNIC software to optimize for the number of peaks, peak positions, and peak widths. Convergence was attained by minimizing the residual between the observed spectrum and the fitted spectrum. Finally, the positions and percentage contributions of the fitted peaks within the spectral region of 1630 to 1700 cm^{-1} were obtained to assign the types of secondary structures and their respective amounts.

The Raman spectra collected on lysozyme during cation exchange purification were analyzed similarly , except that instead of pure water, the Raman spectrum of 20 mM MES buffer with 1 M NaCl at pH 6.2 was used to subtract the background information.

For SERS data analysis, the background spectrum was collected using water. The rest of the data analysis was the same as explained above.

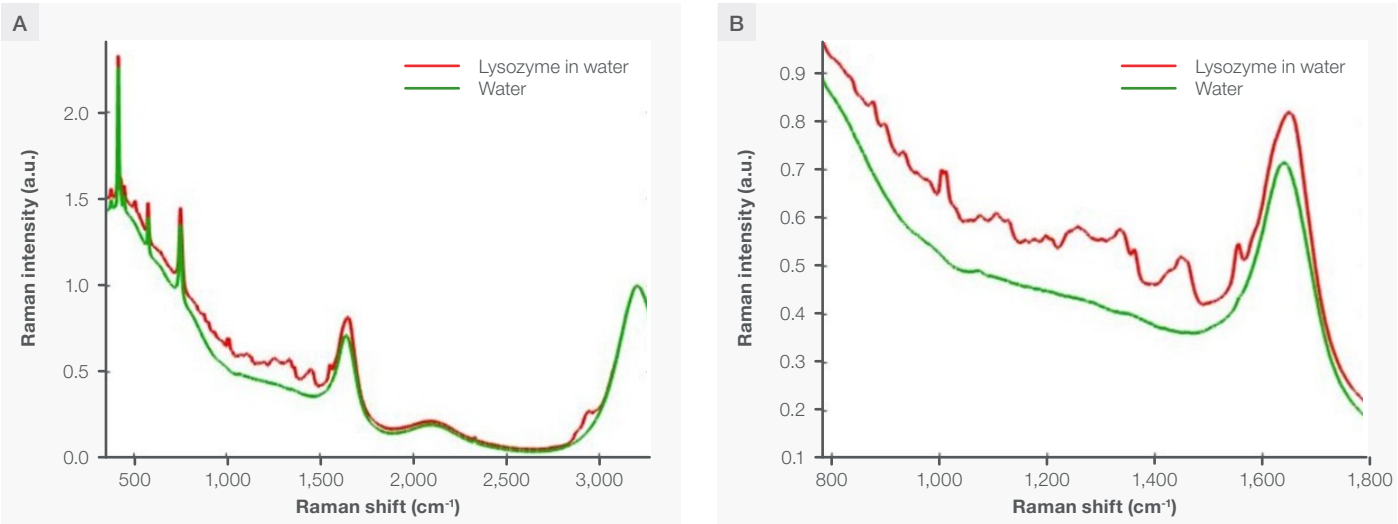


Figure 1. Spectral overlay of 17.5 mg/mL lysozyme in water (red) and water (green) is shown in plot A. The fingerprint region between 800 to 1800 cm^{-1} Raman shift is shown in plot B.

Results and Discussions

Initial PoC study

The spectrum of 17.5 mg/mL lysozyme is shown in Figure 1. In Figure 1A, the spectra are normalized using infinity norm calculated in the region 3000 to 3240 cm^{-1}. Distinct Raman features of lysozymes in the fingerprint region are shown in Figure 1B; these match with reference values reported in the literature.[8] The spectral overlay after removal of baseline for the Raman shift region 1500 to 1800 cm^{-1} is shown in Figure 2. The pure lysozyme spectrum obtained after subtracting water background is shown in Figure 3A. Likewise, the second derivative plot of Figure 3A is shown in Figure 3B which illustrates the presence of four peaks (blue arrow) within the region of 1630 to 1700 cm^{-1}.

The result of peak deconvolution of Figure 3A using Voigt function is shown in Figure 4. The Voigt function was selected based on the recommendation from the literature for aqueous biological samples.[9] The spectral region 1620 to 1750 cm^{-1} of protein is called Amide I region. The peaks in the Amide I region are attributed to the symmetric stretching of the carbonyl functional group in the peptide bond. The carbonyl groups in different secondary structures experience different electronic environments, leading to different vibrational energies that appear as peaks at different positions. Thus, the Amide I region provides information on the secondary structure of proteins.9

The symmetric stretching of the carbonyl group in the peptide bond is both Raman and infrared (IR) active because this vibrational mode is associated with change in polarization and dipole moment during transition, making both Raman and IR spectroscopic techniques suitable for secondary structure analysis. In the literature, offline FTIR, FT-Raman and drop coat deposition Raman (DCDR) have been reported for the study of native protein secondary structures as well as monitoring structural changes during protein degradation, denaturation, aggregation, and chemical modification. However, the possibility of elucidating the protein secondary structure in aqueous phase using in-line Raman has not been reported, which sets the foundation for this study.

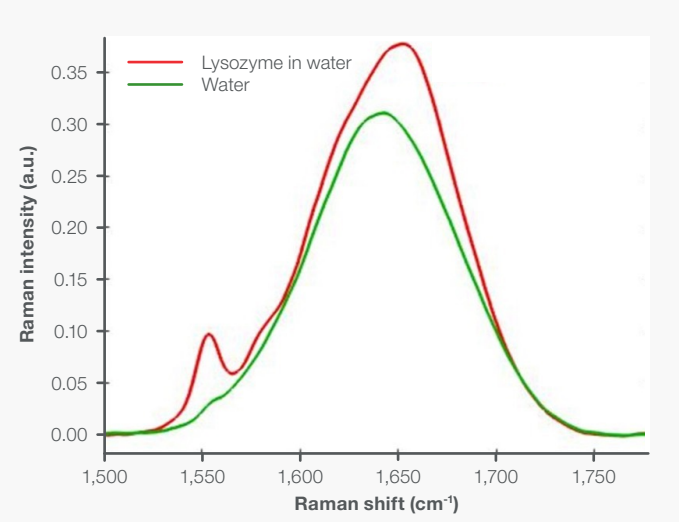


Figure 2. Spectral over of selected region after baseline removal. The lysozyme peak is distinct at approximately 1660 cm^{-1}.

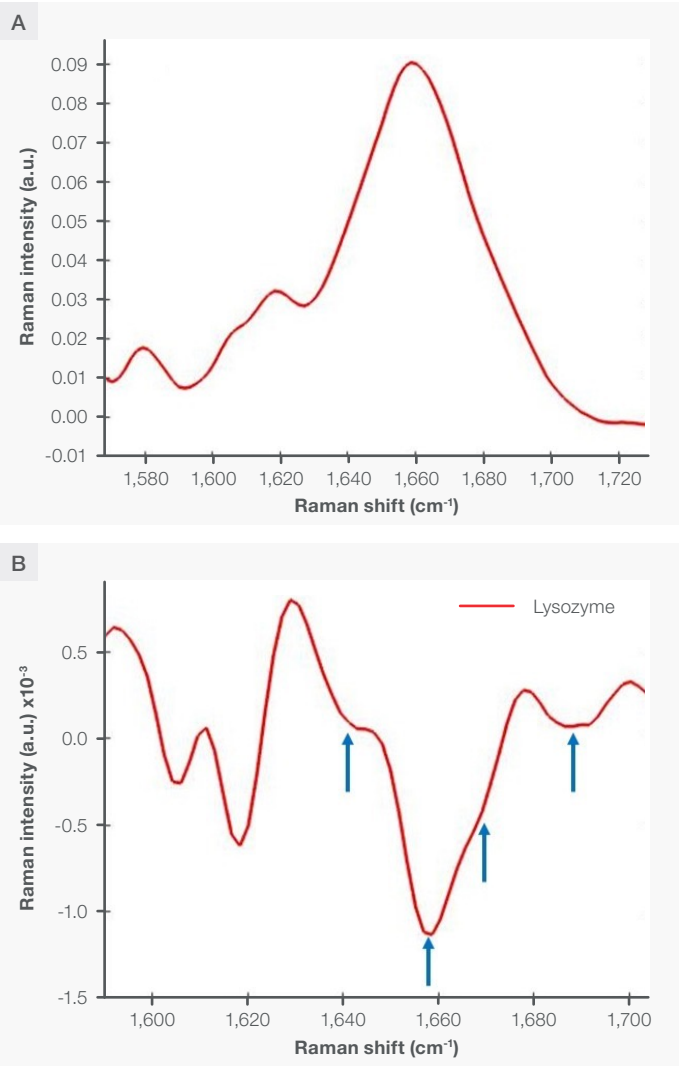


Figure 3. The water background-subtracted lysozyme spectra is shown in plot A and its 2nd derivative is shown in plot B. The blue arrows in the plot B shows four peaks in the spectral region of 1630 to 1700 cm^{-1}.

Secondary Structure	Process Raman		FT-Raman		FT-IR		X-Ray	
	Peak Center	% Area	Peak Center	% Area	Peak Center	% Area	Peak Center	% Area
Solvent Exposed Extended Conformation	1640	24	1637	27	1645	19	NA	19
α-Helix	1657	42	1655	42	1654	40	NA	45
β-Sheet	1672	23	1673	27	1670	20	NA	23
Extended and PPII	1688	11	1685	4	1683	21	NA	13

Table 1. Results from peak convolution from process Raman and its comparison with published result using FT-Raman, FTIR, and X-ray crystallography.

The peak positions and peak contributions in the Amide I region of lysozyme from Figure 4 are summarized in Table 1. The results from process Raman showed that there were four peaks centered at Raman shift of 1640, 1657, 1672 and 1688 cm^{-1} that were assigned to the extended conformation (random coils), α-helices, β-sheet, and extended/PPII (β-turns) respectively. The assignments were adopted based on published work.8 The peak positions and their relative contribution to the Amide I region were in close agreement with the published work with FTIR, FT-Raman, X-ray crystallography, and DCDR. These results clearly demonstrated that the data quality from the MarqMetrix™ All-In-One analyzer provides a platform for the direct analysis of the secondary structure of proteins in the aqueous phase without any sample preparation.

Cation exchange purification

After initial validation of the possibility of in-line protein secondary structure analysis, the strategy was applied to the in-line process data collected during elution step of a lysozyme form column packed with POROS™ 50 HS Strong Cation Exchange Resin. The eluted lysozyme concentration was approximately 25 mg/mL as confirmed by offline analysis. The result for peak deconvolution of Amide I region of lysozyme (1630 to 1700 cm^{-1}) from the in-line process data is shown in Figure 5. A small peak at approximately 1700 cm^{-1} was also observed that was initially assumed to be associated with aggregated lysozyme. However, literature review on Raman based studies on lysozyme aggregation revealed that the peak at 1700 cm^{-1} was not associated with lysozyme. Thus, four peaks centered at 1640, 1656, 1670, and 1684 cm^{-1} were only considered Amide I peaks. The percentage area contributions for these selected peaks, shown in Figure 5, were in close agreement with values shown in Table 1. Note, even if the uncertain 1700 cm^{-1} were taken into consideration, it would have affected the results insignificantly. These results clearly demonstrated that process Raman is a valuable in-line process analytical technology (PAT) that can provide real time insights into protein secondary structure as the downstream process is happening.

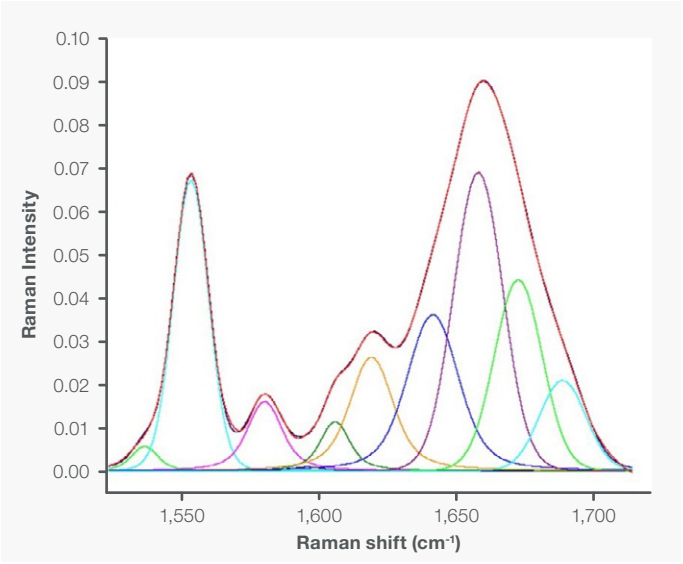
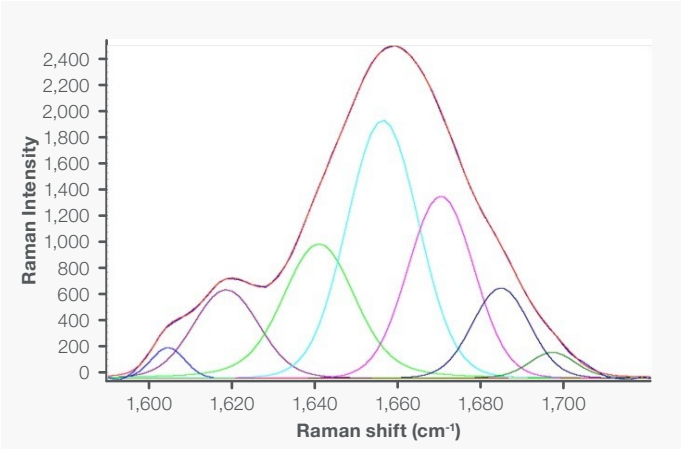


Figure 4. The result of peak deconvolution of lysozyme spectra of Figure 3A is shown. The four peaks in the spectral region between 1620 to 1750 cm^{-1} corresponds to four different secondary structure of lysozyme.

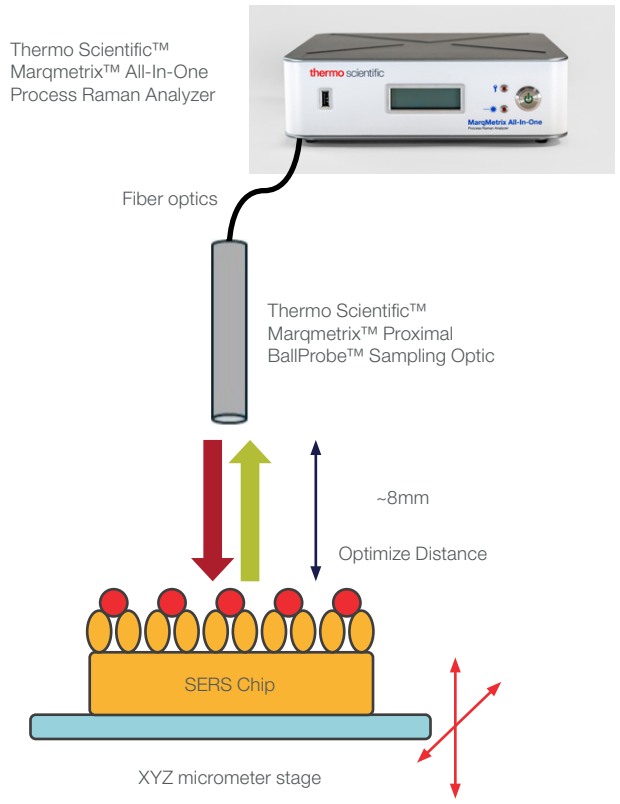


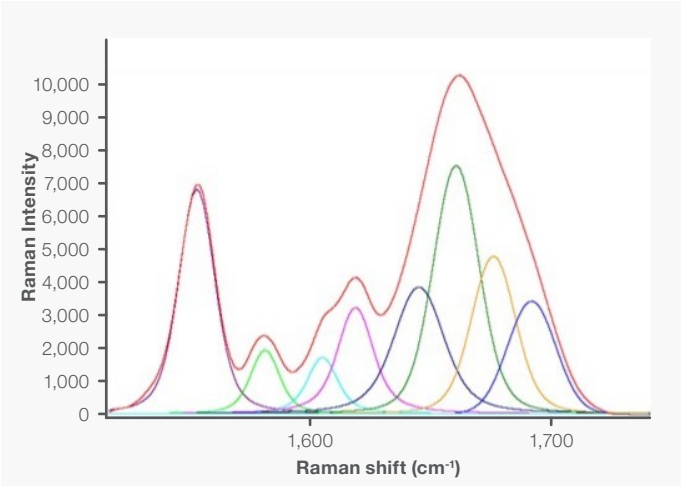
Secondary Structure	Raman Peak Center	%Area
Solvent Exposed Extended Conformation	1640.803	22.59
α-Helix	1656.157	41.93
β-Sheet	1670.16	24.47
Extended and PPII	1684.597	11.01

Figure 5. Showing result of peak deconvolution performed on in-line process data using process Raman.

SERS

SERS is often used as direct or indirect sensor to identify and quantify biological or non-biological analytes due to its high sensitivity. In some cases, the detection limit is close to a single molecule.[10] Several types of SERS exist including colloidal solution, suspended nanoparticle in solution, immobilized nanoparticles on solid surface, nanostructure fabricated on solid surface, or metallic nanowires/foils. Because of high stability, sensitivity, reproducibility of results, low background signal, and homogeneous signal enhancement across the surface, the internally developed SERS substrate was evaluated for this study to analyze the protein secondary structure. The experimental set up and the results for lysozyme secondary structure analysis using SERS substrate is shown in Figure 6. The laser spot size of ~ 500 µm allows sampling from larger area of SERS substrate while the optical fiber provides easy connectivity and sample handling during SERS measurement using the MarqMetrix All-In-One analyzer coupled with the Thermo Scientific™ MarqMetrix™ Proximal BallProbe™ Sampling Optic. A non-contact probe was used to avoid any physical disturbance to the dried protein sample on the SERS substrate and to prevent contamination or degradation that could interfere with the signal. This setup also enabled easier handling of delicate low-concentration samples. The peak deconvoluted results for Amide I of lysozyme from SERS were like that shown in Table 1. Thus, SERS is a viable option to study protein secondary structure when the sample is at low concentration.





Secondary Structure	SERS	
	Peak Center	% Area
Solvent Exposed Extended Conformation	1644	23
α-Helix	1659	41
β-Sheet	1675	24
Extended and PPII	1691	12

Figure 6. Demonstrating the workflow for SERS measurement using the MarqMetrix All-In-One Process Raman analyzer and the output of peak deconvolution for Amide I region of lysozyme.

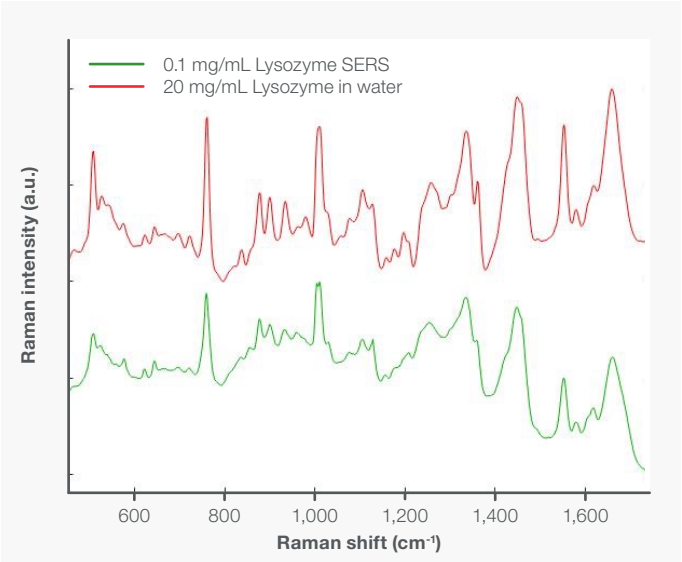


Figure 7. Demonstrating the similarity of Raman spectrum of aqueous 20 mg/mL lysozyme (baseline corrected and water background subtracted) and 0.1 mg/mL lysozyme (without baseline correction) on SERS substrate.

As shown in Figure 7, the Raman spectrum of 20 mg/mL lysozyme (red) and 0.1 mg/mL lysozyme on SERS surface (green) have overall similar peaks profile but minute differences in the spectral regions at 1358 cm^{-1} (CH deformation), 1210 cm^{-1} (aromatic residue), 935 cm^{-1}(C-C stretch), ratio of 508 cm^{-1}, 523 cm^{-1}, 540 cm^{-1} (disulfide region) were also observed. This is expected for larger molecules as not all sections of lysozyme experience homogeneous electro-chemical SERS enhancement effect.[11] Nonetheless, the result shown in Figure 7 demonstrates these SERS substrates are homogeneous and can be used for protein analysis. To test enhancement efficiency of SERS substrate, using the same methodology the data was collected on the non-SERS substrate for 0.1 mg/mL lysozyme but only a weak spectrum was observed (data not shown) under the experimental conditions. Thus, a significant portion of enhancement was attributed to SERS effect, not increase in concentration due to drying as is the case in DCDR.

Another key aspect of SERS measurement is the enhancement of Raman signal relative to fluorescence. The SERS spectra of lysozyme (green) in Figure 7 is shown without baseline removal while aqueous 20 mg/mL lysozyme (red) is plotted after removal of baseline. Thus, SERS can be advantageous for samples that have high fluorescence. Although the details are beyond the scope of this work, a plot is shown in supplementary information that illustrates the enhancement of Raman signals even with high fluorescent background (Figure S1).

Conclusion

In this study, the feasibility of using the Thermo Scientific MarqMetrix All-In-One Process Raman Analyzer for analyzing the secondary structure of lysozyme is demonstrated. The lysozyme secondary structures that were determined were comparable to those obtained using FTIR, FT-Raman, DCDR, and X-ray crystallography. This method's accuracy and the lack of need for sample preparation offer significant advantages for probing protein secondary structure in its native aqueous state. This capability can be leveraged to determine protein secondary structure during real-time downstream processing (e.g., UF/DF), where protein concentrations are relatively high, aiding in the acquisition of high SNR spectra. This ensures the quality of proteins during purification, as published work has shown changes in secondary structure with protein denaturation, aggregation, or degradation. Thus, this capability helps users make actionable decisions to proceed with further downstream processes confidently or, in some cases, scrap the run when quality is compromised. In conjunction with previous work, this study establishes process Raman as a PAT tool for both real-time quantification and quality assessment.

SERS-based protein secondary structure analysis is also demonstrated in a sample with low concentration lysozyme. The protein secondary structures determined from SERS and aqueous solutions were similar, and the accuracy was comparable to FTIR, FT-Raman, DCDR, and X-ray crystallography. The SERS measurement was performed using the MarqMetrix All-In-One Process Raman analyzer, leveraging its features to easily swap probe types and use optical fibers for ease of use in laboratory settings.

References

1. Branden, C. I.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Science: New York, 2012. https://doi.org/10.1201/9781136969898.
2. Wang, H.-W.; Wang, J.-W. How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other. *Protein Sci.* **2017**, *26* (1), 32–39. https://doi.org/10.1002/pro.3022.
3. Pelton, J. T.; McLean, L. R. Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal. Biochem.* **2000**, *277* (2), 167–176. https://doi.org/10.1006/abio.1999.4320.
4. Peters, J.; Jin, C.; Luczak, A.; Lyons, B.; Kalyanaraman, R. Machine Learning Enabled Protein Secondary Structure Characterization Using Drop-Coating Deposition Raman Spectroscopy. *J. Pharm. Biomed. Anal.* **2025**, *259*, 116762. https://doi.org/10.1016/j.jpba.2025.116762.
5. Khadka, N.; Pleitt, K.; Nolasco, M. A Classical Least Squares (CLS) Approach for Protein Quantification in Downstream Processing Using Raman Spectroscopy.
6. Nolasco, M.; Pleitt, K.; Khadka, N. Using a Process Raman Analyzer as an In-Line Tool for Accurate Protein Quantification in Downstream Processes.
7. Nolasco, M.; Pleitt, K.; Khadka, N. Raman-Based Accurate Protein Quantification in a Matrix That Interferes with UV-Vis Measurement.
8. 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.
9. Peters, J.; Park, E.; Kalyanaraman, R.; Luczak, A.; Ganesh, V. Protein Secondary Structure Determination Using Drop Coat Deposition Confocal Raman Spectroscopy. **2016**, *31*, 31–39.
10. Schlücker, S. Surface-Enhanced Raman Spectroscopy: Concepts and Chemical Applications. *Angew. Chem. Int. Ed.* **2014**, *53* (19), 4756–4795. https://doi.org/10.1002/anie.201205748.
11. Pérez-Jiménez, A. I.; Lyu, D.; Lu, Z.; Liu, G.; Ren, B. Surface-Enhanced Raman Spectroscopy: Benefits, Trade-Offs and Future Developments. *Chem. Sci.* **2020**, *11* (18), 4563–4577. https://doi.org/10.1039/D0SC00809E.

Supplementary information

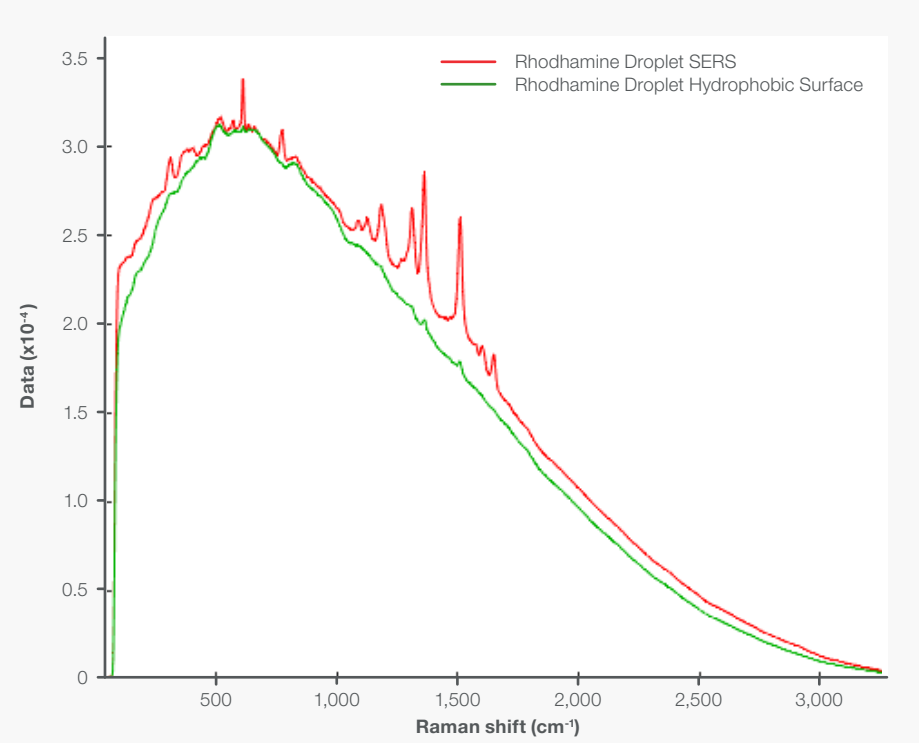


Figure S1. Demonstrating of SERS substrate enhancement of Raman signal even in presence high fluorescence. Both spectra were collected using liquid samples using same acquisition parameters, demonstrating SERS substrate for both liquid and solid samples.

Learn more at **thermofisher.com**