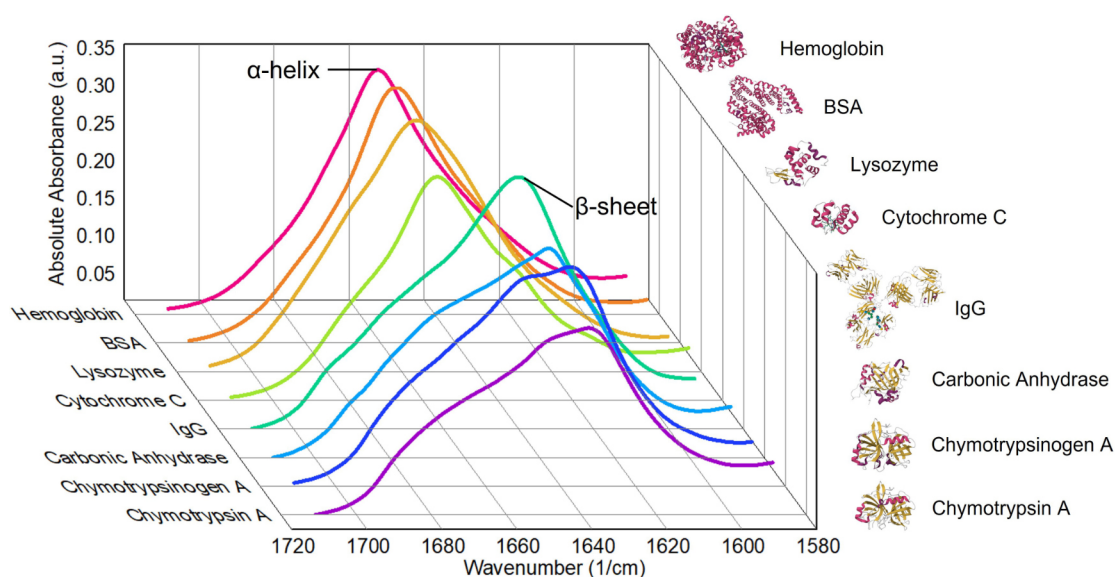


Rapid and Accurate Determination of the Higher Order Structure of a Library of Proteins Using Microfluidic Modulation Spectroscopy

Abstract

Solving the 3-dimensional structure of a protein requires a tremendous amount of time and effort including labor-intensive workflows and complicated data analysis. High-resolution crystal structures are not always a "must" when it comes to protein structural analysis. The primary and secondary structures of proteins are as important as the 3-dimensional structure because most often the early signs of instability and aggregation of proteins are due to mutations in the primary sequence and/or incorrectly folded secondary structures. Here, a library of well-characterized proteins with a variety of conformations is studied using Microfluidic Modulation Spectroscopy (MMS), an automated infrared-based technology that allows rapid and accurate measurement of the secondary structure of proteins. Results showed robust spectral data and higher-order structures that are consistent with the crystal structures of the proteins. These highly reproducible data were used to create additional model proteins in the **delta** analytical software. This app note demonstrates the ability of MMS to generate accurate and reliable structural details of proteins with only a fraction of the time and effort needed for obtaining the crystal structure using techniques such as X-ray crystallography or cryo-EM.



Introduction

In the physiological environment, biomolecules such as proteins and nucleic acids fold into a complex 3-dimensional structure. Making sure the biomolecules have the correct structures is crucial in therapeutic drug production. As structure determines function and activity, a small change in structure such as a re-orientation in the binding pocket could result in a loss of binding recognition. X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (Cryo-EM) are the gold standards to obtain high-resolution structures for proteins. While these techniques provide the highest resolution of the 3-dimensional structures, they suffer from numerous drawbacks such as labor-intensive workflows and complicated data analysis, making them difficult to perform as day-to-day analytical tools.

REDSHIFTBio®

80 Central Street | Boxborough MA | 01719 | info@redshi bio.com

- ☐ Biosimilars
- ☐ mAbs
- ☐ ADCs
- ☐ AAVs
- ☐ Ligand Binding
- ☒ Protein/Peptide Analysis
- ☐ VLPs
- ☐ Nucleic Acid
- ☐ Fusion Proteins
- ☒ Enzyme Analysis

- ☒ Aggregation
- ☒ Quantitation
- ☒ Structure
- ☒ Stability
- ☒ Similarity

Introduction, continued

When it comes to protein structural characterization, all levels of protein structures can be considered. The secondary structure (α -helix, β -sheet, etc.) of proteins contains a tremendous amount of information that is often overlooked by researchers. For example, secondary structure is a key determinant of aggregate formation: the propensity of forming β -sheet in a primary sequence dictates the likelihood of irreversible protein aggregation.¹

It is therefore critical to accurately determine the relative abundance of different secondary structural motifs. The AQS³pro and Apollo, powered by Microfluidic Modulation Spectroscopy (MMS), are automated infrared (IR) spectroscopy technologies for protein secondary structure analysis. Using a quantum cascade laser and a microfluidic flow cell, MMS delivers extremely high-quality data and offers significant improvement in sensitivity, dynamic range, and accuracy for protein analysis compared to conventional FTIR and far-UV CD techniques as demonstrated previously.^{2,3}

In this study, MMS was used to determine the higher-order structure (HOS) of eight common proteins with a variety of secondary structural characteristics. Individual features of each protein were revealed by the high-quality data generated by MMS. The HOS results are compared with three relevant techniques for protein structure determination: FTIR, X-ray crystallography, and AlphaFold.^{4,5} Our results also indicated extremely high reproducibility of measurements for all the proteins studied. The spectral data of these proteins were used as model proteins for samples in their respective classes in the delta analytical software.

Methods

Four α -helix-rich proteins (hemoglobin, BSA, lysozyme, and cytochrome C) and four β -sheet-rich proteins (IgG, carbonic anhydrase, chymotrypsinogen A, and chymotrypsin A) were analyzed in this study. All proteins were obtained from MilliporeSigma in lyophilized powder form and dissolved in appropriate buffers or water shown in Table 1. The proteins are prepared at 10 mg/mL concentration and diluted in a concentration series of 5, 2, 1, 0.5, and 0.1 mg/mL using their respective buffers. Three replicates of each sample solution and the

Table 1. A list of the proteins analyzed in this study with their buffer information.

Secondary Structure Type	Protein Name	Solvent/Buffer
α -helix rich	Hemoglobin	Water
	Bovine serum albumin	PBS at pH 7.4
	Lysozyme	Water
	Cytochrome C	Water
β -sheet rich	Immunoglobulin G	PBS at pH 7.4
	Carbonic anhydrase	Water
	Chymotrypsinogen A	100 mM sodium phosphate at pH 7.4
	Chymotrypsin A	100 mM sodium phosphate at pH 7.4

referencing buffer were injected into the RedShiftBio AQS³pro at a backing pressure of 5 psi and a flow rate of approximately 1 μ L/s. The sample solution and the referencing buffer were modulated at 1 Hz for background subtraction. The differential absorbance between the sample solution and the buffer was measured within the amide I band (1588-1711 cm^{-1}). The spectral data and HOS information were processed and calculated using the RedShiftBio *delta* analytical software integrated into the system.

Results

I: α -Helix Rich Proteins

Understanding the signature peaks for protein α -helix structures:

The amide I band is the most intense absorption region in the IR spectrum for proteins. This absorption is mainly governed by the stretching vibration of the C=O groups in the protein backbone and is found within the mid-IR region between 1600 and 1700 cm^{-1} . Where exactly this C=O absorption peak is in this 100-wavenumber region is dictated by the protein structure. A polypeptide chain is folded into protein secondary structure by forming hydrogen bonding interactions between the backbone atoms via the oxygen on a C=O group and the hydrogen on another backbone N-H group. The peptide backbones thus have different torsion angles and form different hydrogen bond lengths in different secondary structures, i.e., α -helix and β -sheet, thereby causing the C=O group to absorb at different wavenumbers. α -Helices, for example, absorb at around 1656 cm^{-1} .⁶

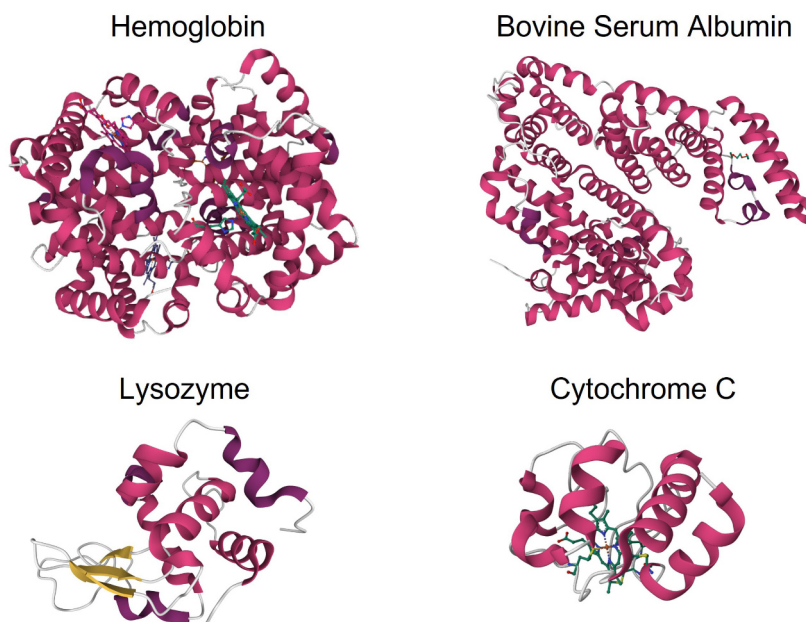


Figure 1. Crystal structures of α -helix-rich proteins: hemoglobin (PDB: 2QSS), BSA (PDB: 3V03), lysozyme (PDB: 1DPX), and cytochrome C (PDB: 1HRC).

Four α -helix-rich proteins with known crystal structures (Figure 1) are studied using MMS. As shown in Figure 1, hemoglobin, BSA, and cytochrome C consist of only α -helix and coil/turn structures, whereas lysozyme is predominantly α -helix, then coil/turn with 3 short strands of β -sheet structures.

These specific structural features, often times obtained from highly resolved crystal structures of the proteins, are accurately determined by MMS. Figure 2 shows the second derivative of the MMS spectra of these four α -helix-rich proteins, highlighting the main peak position in each protein. Sharp peaks in each spectrum appear at 1655–1657 cm^{-1} where α -helices have the highest absorption.⁶ Additionally, a small shoulder next to the main peak in the lysozyme spectra can be observed at around 1640 cm^{-1} . This is attributed to intramolecular β -sheets which have a range of absorption between 1632 to 1642 cm^{-1} . Qualitatively, the results from MMS are in total agreement with the crystal structures of the proteins.⁶

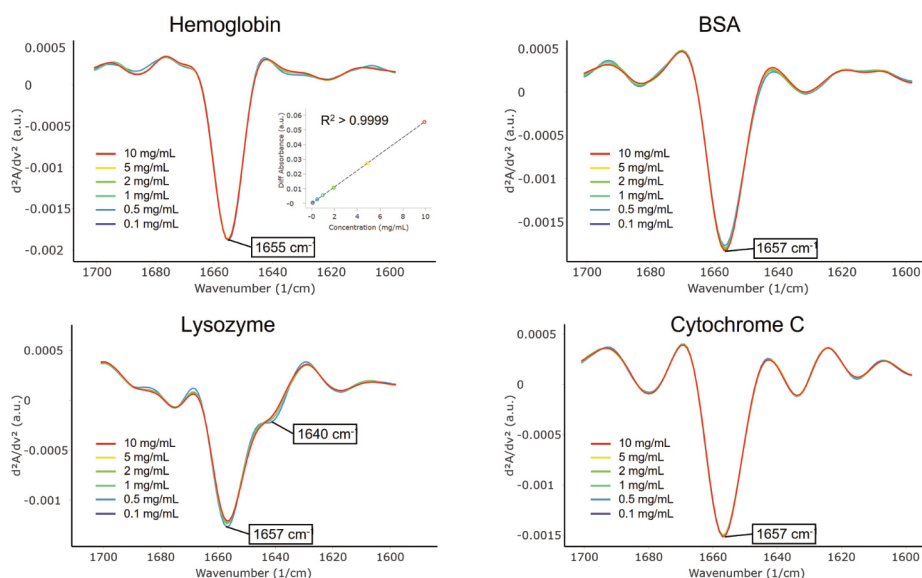


Figure 2. Second derivative spectra of α -helix-rich proteins: hemoglobin, BSA, lysozyme, and cytochrome C. Inset in the hemoglobin spectra shows the quantitation linearity of the concentrations measured, from 0.1 to 10 mg/mL.

Application Note
FEB 2023

A great advantage of MMS is that the *delta* analytical software is already built-in with the system, making data processing a seamless and effortless step after data acquisition. The HOS analysis, being part of the data processing workflow, provides a direct readout of the relative abundance of each of the secondary structural motifs in the entire protein structure. The HOS including α -helix, β -sheet, coil (unord), and turn structures was calculated by Gaussian curve fitting using the inverted and baselined plots of the second derivative spectra.⁷ These quantitative results are compared directly with FTIR,^{8,9} X-ray crystallography,^{8,9} and AlphaFold^{4,5} shown in Figure 3.

It is important to note that while MMS and FTIR determine protein structures in solution, X-ray determines the structures of the proteins in a solid state. In the cases of lysozyme and cytochrome C, both MMS and FTIR measured a higher percentage of β -sheets compared to X-ray and AlphaFold, suggesting that there are more β -sheet structures in these proteins in solution than in crystal form. In addition, the types of buffer and pH can also influence the secondary structure of proteins, as studied previously by MMS [AN-850-0124, AN-850-0125]. Overall, the HOS results agree with one another and have similar patterns across all four platforms.

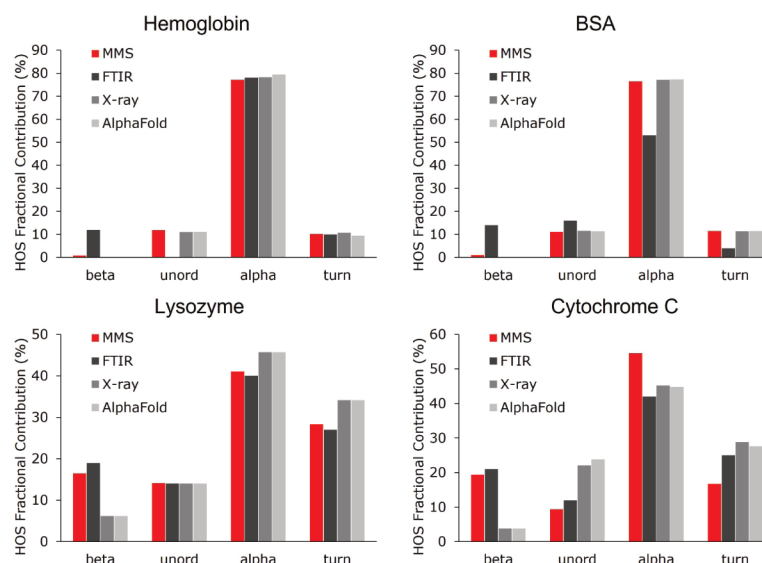


Figure 3. Higher order structure bar graphs showing the relative abundance of the secondary structural motifs for each protein, compared across four different structural analysis platforms: MMS, FTIR, X-ray crystallography, and AlphaFold. MMS data shown used the values from the 10 mg/mL samples.

Results, continued

II. β -Sheet Rich Proteins

Understanding the signature peaks for protein β -sheet structures:

Both α -helix and β -sheet structures are formed by hydrogen bonding interactions between the protein backbone. However, they differ by the types of hydrogen bonding which gives rise to different oscillation frequencies of their backbone C=O bonds. Consequently, α -helix and β -sheet can be confidently distinguished by the shape and position of the associated absorption bands. While α -helix structures are typically very robust and have a narrow absorption band, β -sheet structures can come in two different types and their absorption bands are wider in general. The intramolecular β -sheet is the native structure that exists in proteins, and it absorbs at $1632\text{--}1642\text{ cm}^{-1}$ as mentioned in the previous section. When proteins aggregate, resulting from unfolding, the native β -sheets can form intermolecular interactions and eventually tightly bound β -sheets such as β -amyloids. These intermolecular β -sheets absorb at $1618\text{--}1624\text{ cm}^{-1}$ and $1695\text{--}1700\text{ cm}^{-1}$.⁶

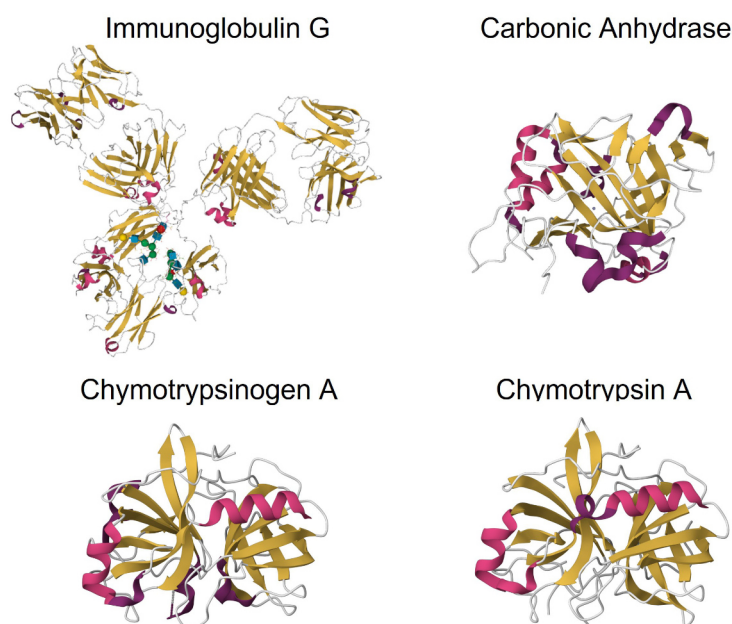


Figure 4. Crystal structures of β -sheet-rich proteins: IgG (PDB: 5DK3), carbonic anhydrase (PDB: 1V9E), chymotrypsinogen A (PDB: 2CGA), and chymotrypsin A (PDB: 4CHA).

Four β -sheet-rich proteins with known crystal structures (Figure 4) are studied using MMS. As shown in Figure 4, all structures consist of predominantly β -sheet with some α -helix and coil/turn structures. MMS was used to detect and analyze these secondary structures. Figure 5 shows the second derivative spectra of these four proteins, with the main peaks marked in each plot. Sharp peaks in each spectrum appear at 1635-1639 cm^{-1} , indicating that the main secondary structure in each protein is intramolecular β -sheet. For both chymotrypsinogen A and chymotrypsin A, a minor peak arises around 1650 cm^{-1} . This peak is assigned to the coil structure in these proteins.⁶ Interestingly, despite the structural similarity between these two proteins (chymotrypsinogen is the inactive precursor of chymotrypsin), their spectral differences are noticeable (Figure 5). An in-depth study of the structural differences between chymotrypsinogen A and chymotrypsin A and their relation to the activities will be featured in an upcoming app note.

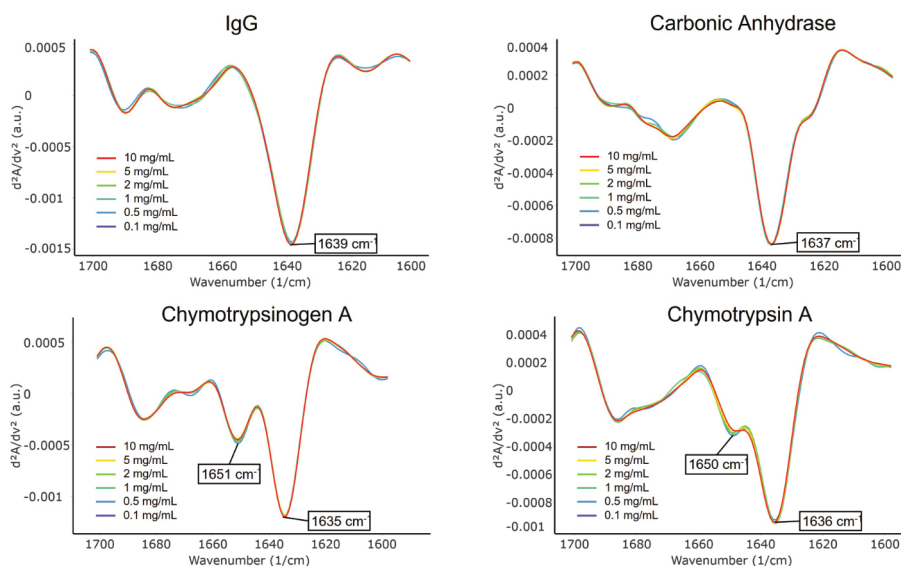


Figure 5. Second derivative spectra of β -sheet-rich proteins: IgG, carbonic anhydrase, chymotrypsinogen A, and chymotrypsin A.

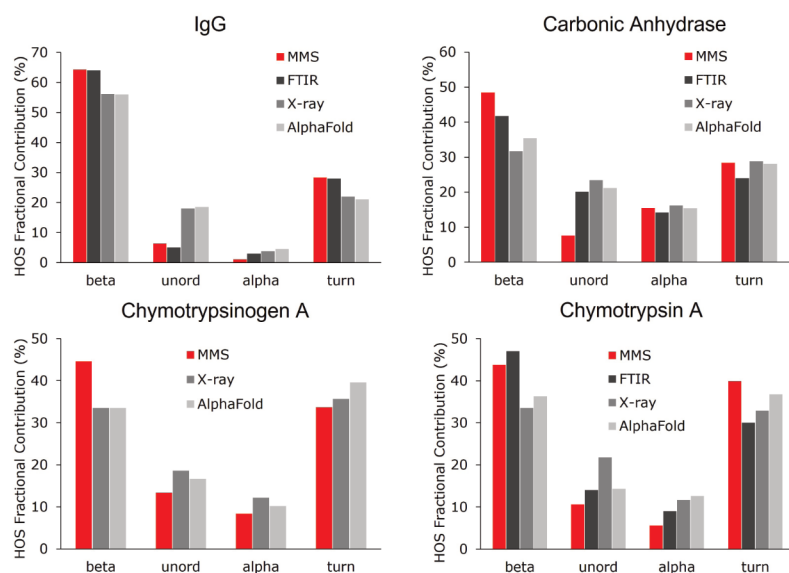


Figure 6. Higher order structure bar graphs showing the relative abundance of the secondary structural motifs for each protein, compared across four different structural analysis platforms: MMS, FTIR, X-ray crystallography, and AlphaFold. MMS data shown used the values from the 10 mg/mL samples. Note: there is no FTIR data available for chymotrypsinogen A.

The HOS of the β -sheet-rich proteins is compared directly with FTIR, X-ray crystallography, and AlphaFold (Figure 6). Overall, the HOS results obtained by MMS agree with those obtained by the other platforms. There are small differences across the four platforms when comparing the individual secondary structural motifs. For example, the β -sheets in IgG (polyclonal) were estimated to be 64% using MMS and FTIR, whereas the percentage dropped to around 55% using X-ray and AlphaFold. Similar trends in the abundance of the β -sheets can be observed in other proteins. These differences are likely due to the fact that MMS and FTIR determine protein structures in solution while X-ray determines the structures of the proteins in solid state. A previous study revealed that protein conformations with more hydrophobic amino acids are more similar in crystal and solution-based forms than those with more hydrophilic amino acids.¹⁰

III. Data reproducibility

MMS utilized a microfluidic cell that can modulate the sample in solution and the referencing buffer every second to allow real-time buffer subtraction. As a result, all the spectra and HOS calculations were obtained based on highly reproducible data. The reproducibility of measurements for each concentration in each protein is shown in Figure 7. Reproducibility is calculated using the area of overlap between each replicate spectrum and the averaged spectrum as previously described.² For all the proteins studied, samples at 1 mg/mL concentration consistently reached at least 98% reproducibility.

Application Note
FEB 2023

Since the lower concentration limit of detection for MMS is 0.1 mg/mL, there is a greater variability of reproducibility at this concentration. Generally speaking, higher sample concentration results in higher signal-to-noise ratio and hence higher reproducibility. Samples at 10 mg/mL reached at least 99.8% reproducibility. The extremely high reproducibility demonstrates the robustness of MMS measurements and brings confidence to seeing actual structural changes in comparability studies

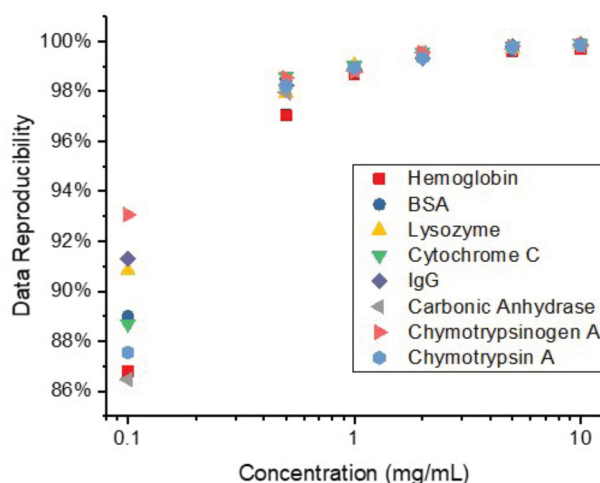


Figure 7. Reproducibility of measurements at different concentrations for each protein.

Conclusions

A library of common proteins with a variety of secondary structural characteristics was studied using MMS, and the HOS was determined and compared to other structural characterization or prediction tools: FTIR, X-ray crystallography, and AlphaFold. Overall, the HOS calculated using MMS is in agreement with the other techniques. Some differences in the amount of β -sheet or α -helix structures can be noticed between solution-based MMS, FTIR, and solid-based X-ray crystallography, as expected and demonstrated previously.¹⁰ Buffer conditions, pH, and protein concentrations also play a potential role in such differences between these techniques. Furthermore, the robustness of the spectral data is validated by the reproducibility of measurements, with >99.8% reproducibility in all the 10 mg/mL samples. In the process, we have expanded our library of model proteins in the delta analytical software using the spectral data in this study. This addition will increase the relevance and accuracy of the processed data by enabling the use of appropriate model proteins to analyze samples with unknown structures.

CONTRIBUTOR(S):

Richard Huang, PhD

References

- (1) Uversky, V. N.; Fink, A. L. Protein Misfolding, Aggregation, and Conformational Diseases. *In Protein Reviews*; Springer, 2006; Vol. 4.
- (2) Kendrick, B. S.; Gabrielson, J. P.; Solsberg, C. W.; Ma, E.; Wang, L. Determining Spectroscopic Quantitation Limits for Misfolded Structures. *J Pharm Sci* 2020, 109 (1), 933–936. <https://doi.org/10.1016/j.xphs.2019.09.004>.
- (3) Liu, L. L.; Wang, L.; Zonderman, J.; Rouse, J. C.; Kim, H. Y. Automated, High-Throughput Infrared Spectroscopy for Secondary Structure Analysis of Protein Biopharmaceuticals. *J Pharm Sci* 2020, 109 (10), 3223–3230. <https://doi.org/10.1016/j.xphs.2020.07.030>.
- (4) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 2021, 596 (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (5) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Židek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res* 2022, 50 (D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- (6) Dong, A.; Huang, P.; Caughey, W. S. Protein Secondary Structures in Water from Second-Derivative Amide I Infrared Spectra; 1990; Vol. 29. <https://pubs.acs.org/sharingguidelines>.
- (7) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by Deconvoluted FTIR Spectra. *Biopolymers* 1986, 25 (3), 469–487. <https://doi.org/10.1002/bip.360250307>.
- (8) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining Information about Protein Secondary Structures in Aqueous Solution Using Fourier Transform IR Spectroscopy. *Nat Protoc* 2015, 10 (3), 382–396. <https://doi.org/10.1038/nprot.2015.024>.
- (9) Kong, J.; Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim Biophys Sin (Shanghai)* 2007, 39 (8), 549–559. <https://doi.org/10.1111/j.1745-7270.2007.00320.x>.
- (10) Sikic, K.; Tomic, S.; Carugo, O. Systematic Comparison of Crystal and NMR Protein Structures Deposited in the Protein Data Bank. *Open Biochem J* 2010, 4, 83–95.