Electronic Supplementary information:

Raman spectroscopic analysis of High Molecular Weight Proteins in solution– considerations for sample analysis and data pre-processing

Drishya Rajan Parachalil[a,b]*, Brenda Brankin[c] , Jennifer McIntyre[a], and Hugh J Byrne[a]

FOCAS Research Institute, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland

School of Physics and Optometric & Clinical Sciences, Kevin Street, Dublin 8, Ireland

**Corresponding Author: drishyarajan.parachalil@mydit.ie*

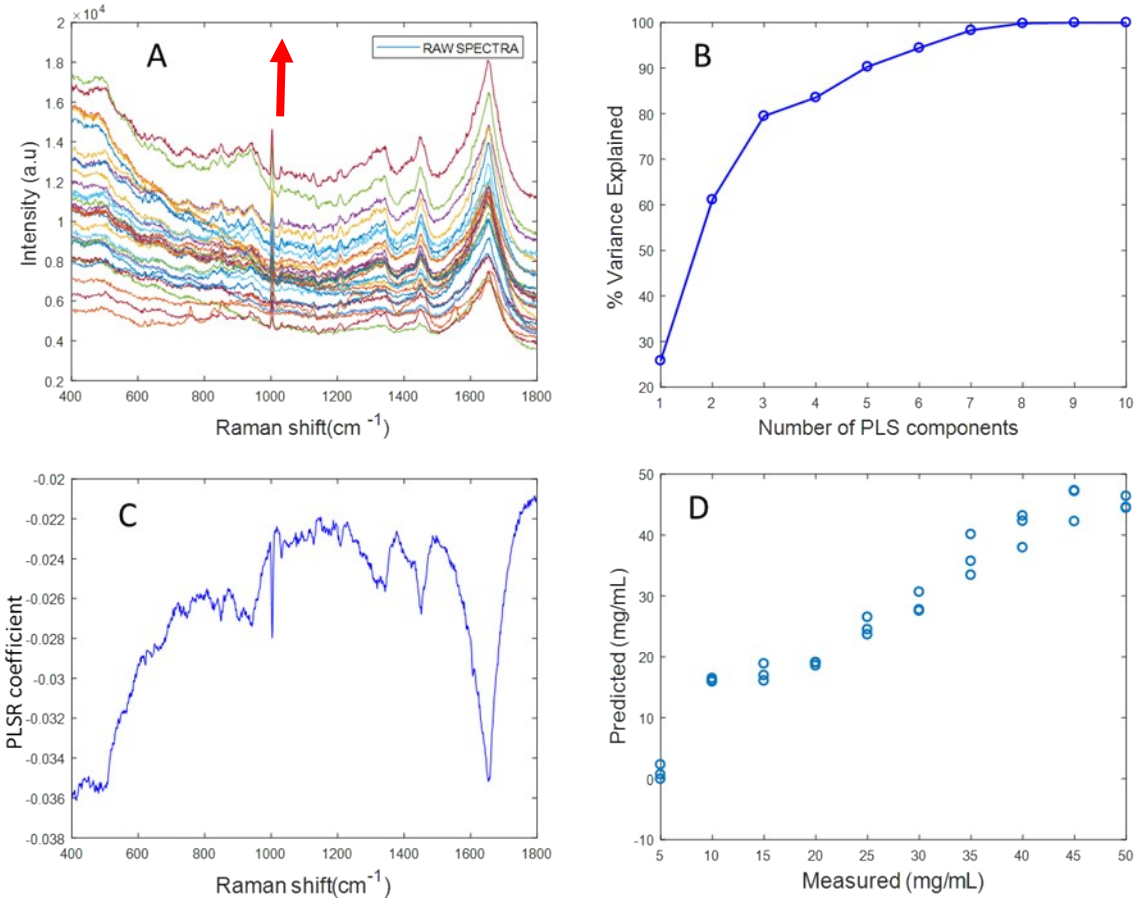

Fig. S1. A: Raw Raman spectra of varying concentrations of albumin in simulated plasma (5mg/mL – 50mg/mL). The arrow indicates the order of increase in concentration, B: Percent variance explained by the latent variables, C: PLSR component showing the inverse albumin features, and D: Linear predictive model built from the PLSR analysis

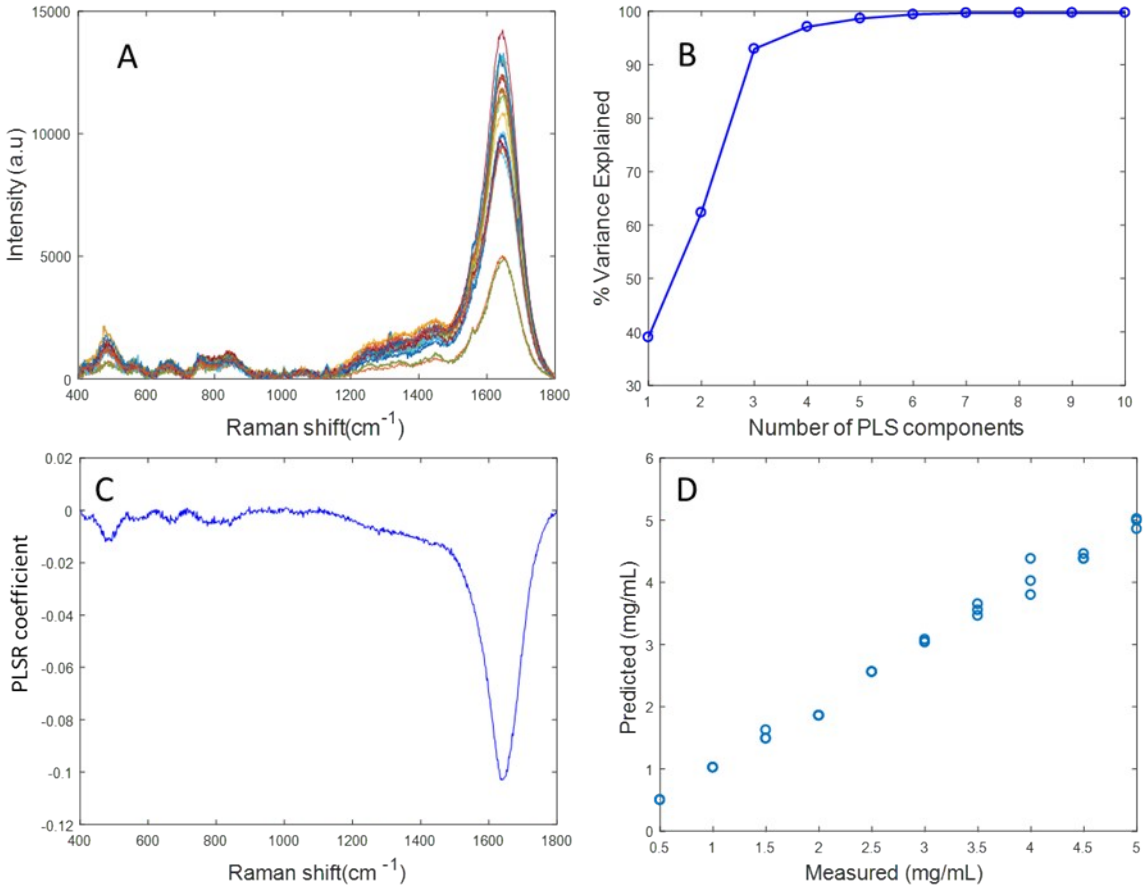

Fig. S2. A: Raman spectra of varying concentrations of fibrinogen in distilled water corrected using the "rubberband" method (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient plotted from the data set shows negative peaks and no fibrinogen features, D: Linear predictive model built from the PLSR analysis

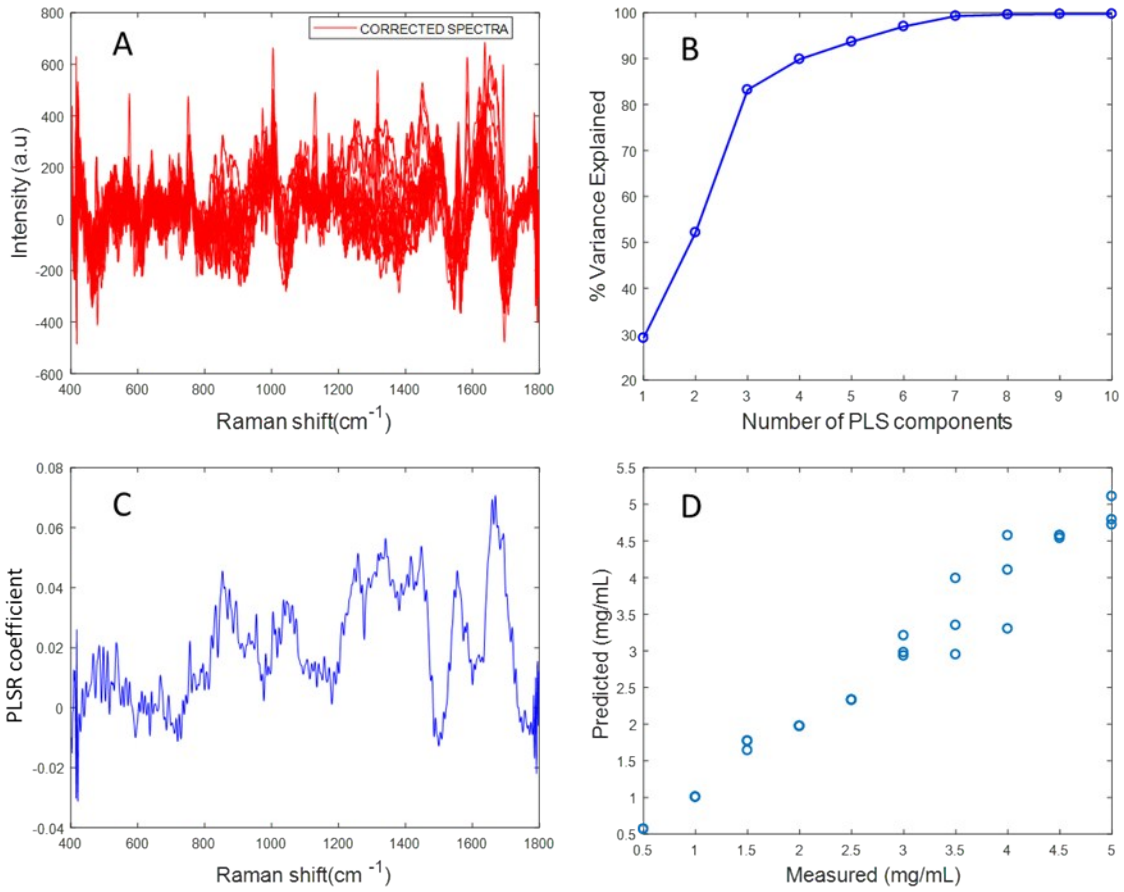

Fig. S3. A**:** Spectra corrected by EMSC algorithm of varying concentration of fibrinogen (0.5mg/mL – 5mg/mL), B: Percent variance explained by the latent variables, C: PLSR coefficient showing strong fibrinogen features, D: Linear predictive model built from the PLSR analysis.

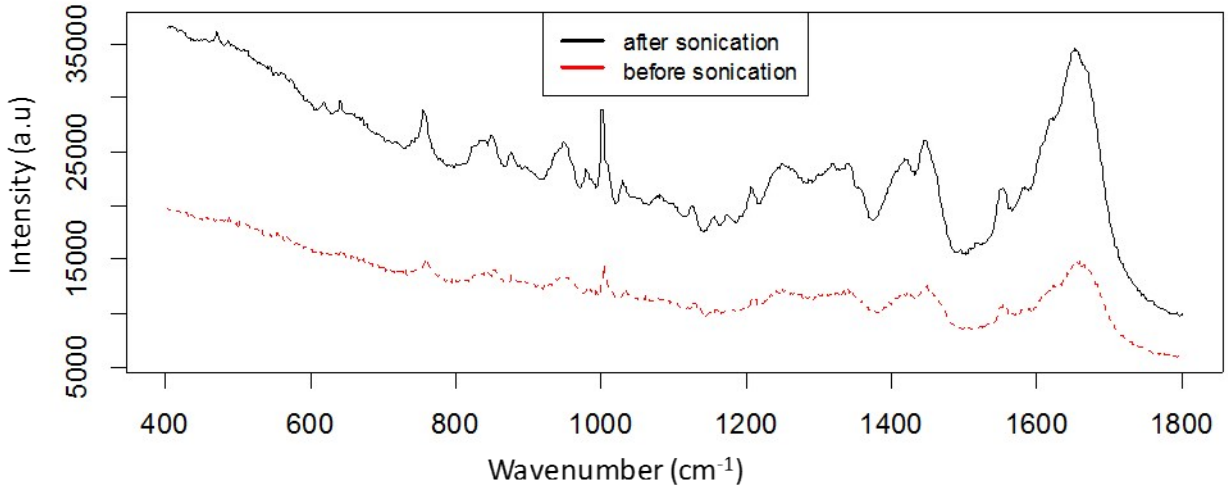

Fig. S4. Raman spectra of fibrinogen stock recorded before (Black) and after (Red) sonication. Sonication helped to increase the overall intensity of the Raman signal by increasing the solubility of the protein.

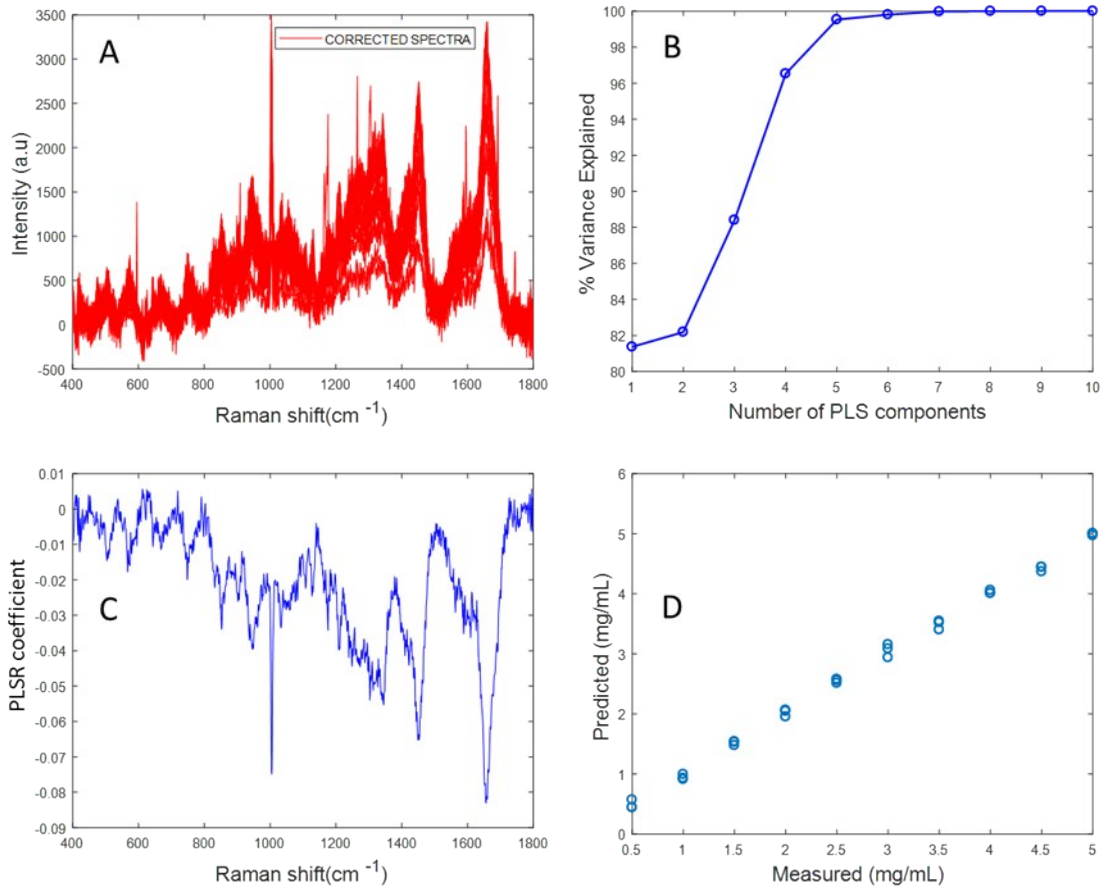

Fig. S5. A: EMSC corrected data of varying concentrations of fibrinogen (0.5mg/mL to 5mg/mL), B: percent variance explained by the latent variables, C: PLSR coefficient showing

the inverse peaks of albumin (1089cm^{-1} and 1102cm^{-1}), and D: The predictive model built from the dataset

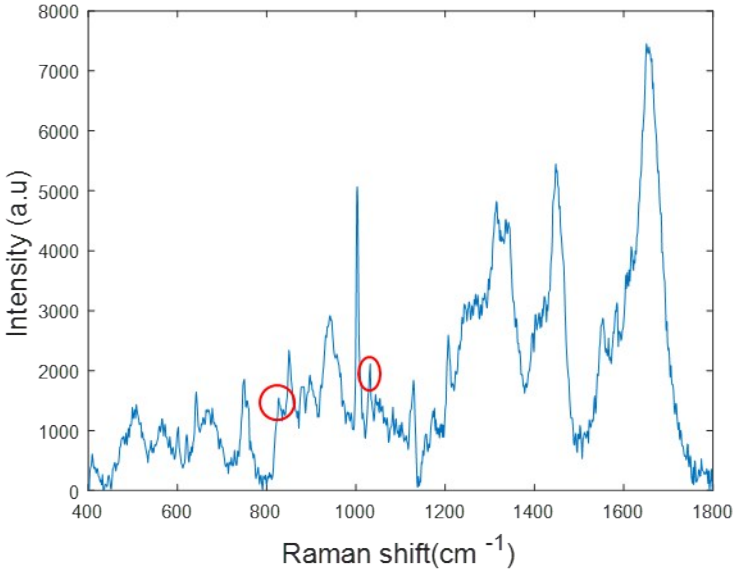

Fig. S6. Raman spectrum of the concentrate obtained after ultracentrifugation of simulated plasma using 100KDa filters showing albumin features (899cm^{-1} and 1102cm^{-1})

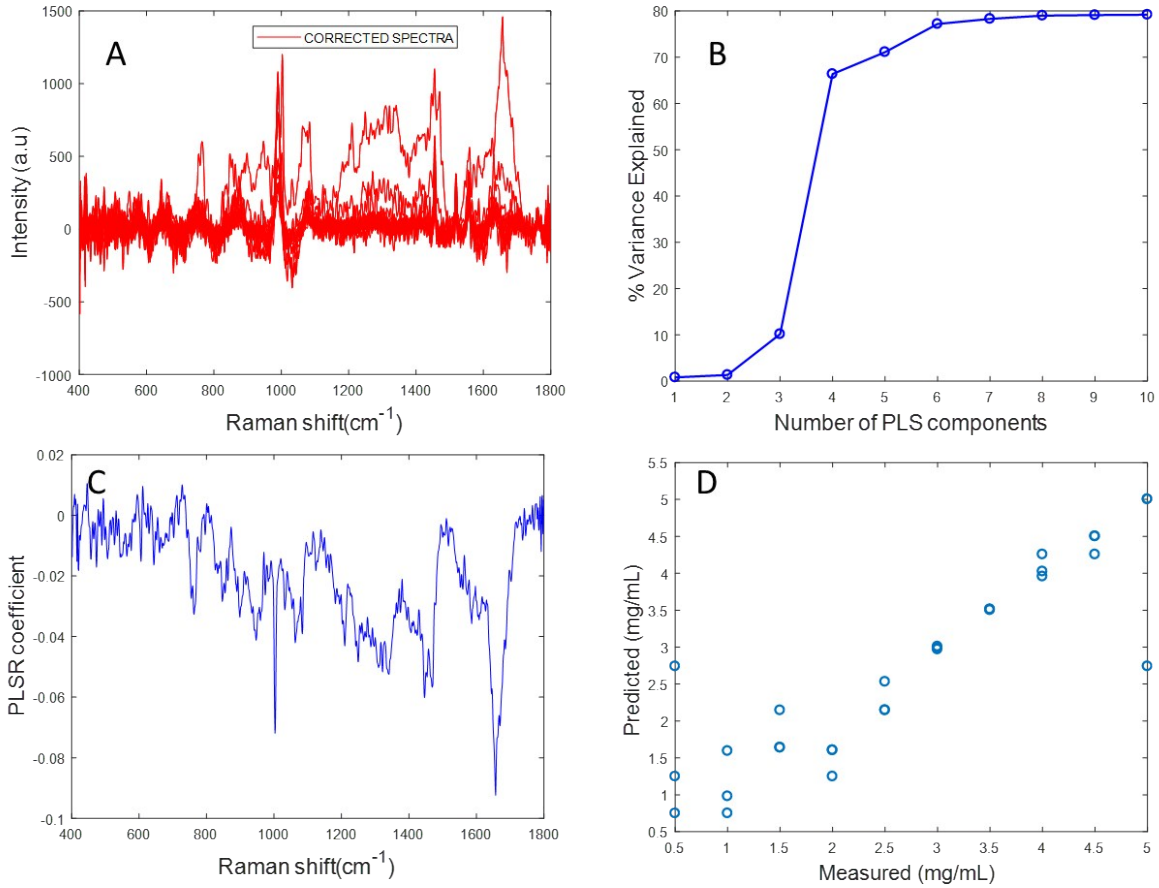

Fig. S7. A: EMSC corrected spectra of varying concentrations of fibrinogen separated by ion exchange chromatography (0.5mg/mL to 5mg/mL), in the absence of sonication B: percent variance explained by the latent variables, C: PLSR coefficient showing inverse fibrinogen peak, and D: The predictive model built from the dataset