

006455

# Talanta

Reproducibility of the nonlinear regression fit to the FTIR  
spectra of lysozyme<sup>1</sup>

Thomas F. Kumosinski \*, Joseph J. Unruh, Harold F. Farrell Jr.

*U.S. Department of Agriculture, ARS, Eastern Regional Research Center, 600 East Mermaid Lane, Wyndmoor, PA 19038, USA*

# Reproducibility of the nonlinear regression fit to the FTIR spectra of lysozyme<sup>1</sup>

Thomas F. Kumosinski \*, Joseph J. Unruh, Harold F. Farrell Jr.

*U.S. Department of Agriculture, ARS, Eastern Regional Research Center, 600 East Mermaid Lane, Wyndmoor, PA 19038, USA*

Received 25 September 1996; received in revised form 22 January 1997

---

## Abstract

Controversy exists concerning the influence of experimental artifacts on the number of component FTIR vibrational bands which may be resolved from the amide I and II envelopes of proteins in water. Whether these bands represent unique populations of vibrating protein groups in a particular global 2° structure or whether the bands are due to instrumental and environmental fluctuations has been addressed, T.F. Kumosinski and J.J. Unruh, *Talanta*, 43 (1996) 199–219. The repeatability of the methodology and the apparent uniqueness of the nonlinear regression fits are addressed in this study. We obtained a series of the spectra of lysozyme, and carried-out nonlinear regression analysis of each spectrum. Coefficients of variation (COV) were calculated for the wavenumber and area values of assigned component peaks obtained. Low COVs obtained attest to the precision of the methodology and the apparent uniqueness of the nonlinear regression fits. This methodology for acquisition and analysis of protein FTIR spectra yields results with good precision. © 1997 Elsevier Science B.V.

**Keywords:** FTIR; Fourier transform infra red spectroscopy; Lysozyme; Precision; Proteins; Protein secondary structure

---

## 1. Introduction

FTIR analysis of proteins and peptides as a method for estimation of their global secondary structure has become a valuable tool for protein biochemists [1,2]. While this methodology is used more frequently, controversy exists in the areas of how to carry out the experiments and analyze the

results. Whether to obtain the spectra in the real-world environment of H<sub>2</sub>O with a difficult subtraction routine, or in the potentially structure changing environment of D<sub>2</sub>O [2,3] with an easier subtraction routine is still debated. The best method for analysis of the results is also in debate. Many methods and combinations for analysis are available to choose from; these include: nonlinear regression analysis; second derivative analysis; Fourier self-deconvolution; and basis-set factor analysis. The pros and cons of the above considerations have been discussed elsewhere [4,5].

---

\* Corresponding author. E-mail: tkumosinski@arserrc.gov/junruh@arserrc.gov

<sup>1</sup> Reference to a brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

Recently we reported an FTIR methodology with which we obtained the spectra of 14 proteins in H<sub>2</sub>O, analyzed the data using the second derivative and nonlinear regression methods, and compared the results with the known X-ray crystallographic data [1]. Differences of 2–4% for the periodic structures of helix, extended, irregular and turn were observed between the FTIR and X-ray results. Thus the accuracy of the method is firmly founded. An additional consideration would be the reproducibility (overall precision) of the spectral acquisition and the uniqueness of the nonlinear regression analysis. Here we present a validation study showing the reproducibility and uniqueness of the results using our reported methodology [1] with the well characterized protein lysozyme.

## 2. Experimental

### 2.1. Materials and solutions

Water was Type I grade obtained from the NANOPURE system (Barnstead, Dubuque, IA). Lysozyme (chicken egg white) and PIPES (potassium salt) were from Sigma (St. Louis, MO). Any other chemicals were of reagent grade or better. The CaF<sub>2</sub> windows and 6 µm Mylar<sup>®</sup> spacer were from SPECAC (Smyrna, GA). The buffer solution was 25 mM PIPES (potassium salt) with pH adjusted to 6.65 using concentrated HCl. After pH adjustment the buffer was filtered through a 0.22 µm Nylon filter (Sigma), and stored at 5°C. Two Lysozyme solutions were made—one at 3% (w/w) and one at 1.5% (w/w)—in the PIPES buffer, then filtered through 0.22 µm DVDF filters (Sigma) and stored at 5°C. The concentration of the lysozyme solutions were verified by UV spectrometry ( $\epsilon_{280} = 36\,000$ ) [6] on a Beckman DU-7 (Fullerton, CA) instrument.

### 2.2. Infrared measurements

The procedures and equipment were as previously reported—collecting 4096 double-sided interferograms (each with 16 384 data points), which were co-added, phase-corrected, apodized (Happ-Genzel function), and fast-Fourier transformed [1].

The following minor exceptions to the methodology were made: a 6 µm Mylar<sup>®</sup> spacer was used in place of the 12 µm Teflon spacer; the sample cell's temperature was controlled at 25°C using SPE-CAC's electrical heating jacket (part # 20707, SPE-CAC); and the lysozyme solution was 3% (w/w) in 25 mM PIPES·K, pH = 6.65 in place of 4% in an imidazole buffer. To test repeatability, a spectrum was acquired each day for 5 days with the sample cell being disassembled and cleaned daily.

### 2.3. Data analysis

The buffer subtraction procedure used to obtain the protein spectra has been reported elsewhere [1]. Briefly, after obtaining the absorbance spectra for the protein and buffer solutions the protein spectra (amide I and II region) were obtained by subtracting the buffer spectra from the respective protein solution spectra in the 2000–1350 cm<sup>-1</sup> region. Subtractions were performed interactively using the subtraction function in the Sx software of the Nicolet 660 data system. The scaling factor (FCR) was individualized during each subtraction by adjusting the FCR parameter value until the region from 2000 to 1700 cm<sup>-1</sup> was as flat as possible. Next, the water vapor absorption was subtracted from the protein spectra using the 2nd derivative method [7].

The data analysis procedure of our previously reported methodology was changed slightly [1]. Here, the Fourier-self deconvolution spectra were not calculated because the choice of values for the parameters is controversial and the results thus questionable. We found more acceptable analyses could be obtained using only the second derivatives and the nonlinear regression fits of the original spectra. As previously reported, the starting wavenumbers for the ABACUS [8]<sup>2</sup> curve fitting program were chosen from the second derivative

<sup>2</sup> Version F.1 is available on the internet from QCPE (their program number 652) or directly from ERRC. Persons desiring the program may use FTP, connect to 'ceres.arserrc.gov', and use the account 'anonymous' which requires no password. They would then select directory 'abacusf' and download all files. Installation instructions and user documentation are also provided.

of the original spectrum. To increase the probability that the resulting fit was not a local minimum and the unique fit to the data, the original spectrum was fit twice, once starting from below the experimental values (where  $Y_{\text{exp}} - Y_{\text{theoretical}}$  is positive) and once starting from above the experimental values (where  $Y_{\text{exp}} - Y_{\text{theoretical}}$  is negative). Convergence to the same model from both starting positions was one criterion for an acceptable fit. The starting peak heights for the 'below fit' were chosen so that the composite sum was at least 25% below the original spectra absorbance plot, and the starting peak heights for the 'above fit' were similarly chosen to be 25% above the original spectra absorbance plot. The starting half-width at half-height was three for all peaks. While just the peak heights were adjusted to obtain two different starting points, it should be emphasized that all peak parameters (height, half-width at half-height, and wavenumber) as well as the baseline were permitted to float during the nonlinear regression analysis. A more detailed discussion of the nonlinear regression analysis can be found in our previous report [1].

Three criteria were required to be met before the results of the curve-fitting were considered acceptable. First, the lowest RMS (root-mean-square) value obtainable had to be reached. Acceptable RMS values were usually of magnitude  $10^{-5}$  for spectra with maximum absorbance values in the 0.05–0.09 range. Second, the below and above fits should converge to the same model. Third, the second derivative of the composite sums should overlies the second derivative of the original spectra. Only when all three of the above criteria were met was the curve-fitting result considered acceptable.

$\chi^2$  instead of the RMS could also be evaluated as a means to determine the best nonlinear fit, since it includes the effects of adding more parameters to the model. For our ABACUS program,  $\chi^2$  and the RMS are equivalent when the model has a fixed number of peaks as it does in our case for any individual fit of a model to the data. In the case where more or less peaks may be needed in the model, we use an  $F$ -test to statistically evaluate the change in the number of parameters (i.e., the number of peaks). This part

of the methodology is discussed in more detail in our previous report [1].

### 3. Results and discussion

#### 3.1. Spectral acquisition

Fig. 1 shows the spectra (amide I and II region) obtained on each of 5 different days for lysozyme. In all but one case the spectra are nearly identical. The one exception sample was prepared at half the concentration. This spectrum was included in the curve-fitting analysis to consider variance with concentration. With the exception of the 1.5% lysozyme spectra the upper four spectra nearly overlies each other, except in the  $1500\text{--}1400\text{ cm}^{-1}$  region. We can not explain this but have not noticed differences in this region to effect the nonlinear regression analysis of the  $1700\text{--}1500\text{ cm}^{-1}$  region. We have noticed though that small changes in the FCR during buffer subtraction have a greater effect in this region than the  $1700\text{--}1500\text{ cm}^{-1}$  region.

During the 5 days of the study, the ambient temperature of the analysis room changed slightly ( $68\text{--}72^\circ\text{F}$ ) and the humidity changed greatly ( $27\text{--}56\%$  relative humidity). As can be observed these environmental changes had little effect on the obtained spectra, and the presence of an  $\text{H}_2\text{O}$  vapor spectra are not observed in the final protein

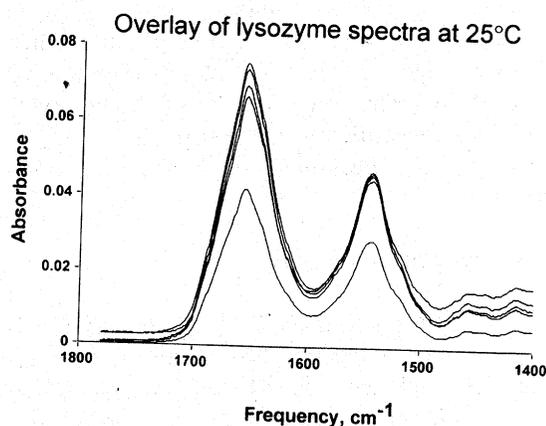


Fig. 1. Overlay plot of lysozyme spectra obtained as described in text.

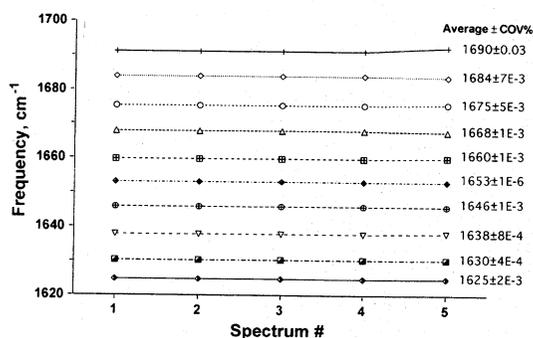


Fig. 2. Graph of the wavenumbers obtained from the nonlinear regression analysis of the spectra in Fig. 1 versus the day of acquisition.

spectra obtained. Daily cleaning and reassembly of the cell showed a variation in the pathlength to be approximately 1.5% Coefficient of variation (COV).

### 3.2. Curve-fitting—nonlinear regression analysis

Two points of evaluation were considered to determine if the curve-fitting procedure leads to repetitive results. First, if the resultant wavenumbers for the component peaks from the fit turned out to be the same for each of the five spectra; and second, if the resultant areas for the component peaks turned out to be the same for each of the five spectra.

Fig. 2 demonstrates the first case. There the wavenumbers found for the amide I of the spectra from each of the 5 days are plotted versus the day number. Nearly flat horizontal lines are obtained. This shows the repeatability of the curve-fitting procedure to obtain the same wavenumbers for different spectra of the same sample. Beside each plotted line is listed the average wavenumber  $\pm$  COV%. The COV for the wavenumbers varies from 0.000002 to 0.03%—indicating very precise results for the wavenumbers. Repeatable wavenumbers show that defining the structural elements in the model for the secondary structure is precise for lysozyme, and will probably be precise for most other proteins analyzed. A more stringent test for repeatability of analysis will be found by comparing the areas for each wavenumber's component peaks.

Table 1 lists the areas of the component peaks obtained for each day's analysis. The standard deviations are good, an average COV of 7.86% (range 1.78–26.71%) is obtained for days 1–5. This shows acceptable repeatability. The greatest variation occurs in the day 5 results, where all the data for this day approach being considered outliers. Day 5 is the sample prepared at 1.5%, to test the effect of concentration on the results. If the results of day 5 are not included, the averages obtained only change slightly (last column in Table 1), while the average COV changes to 2.39% (range 1.36–4.97%). Eliminating the day 5 results greatly improves the precision of the methodology. Our first report [1] concerned itself with the accuracy of this methodology when comparing the results to X-ray crystallographic data. In that report we only analyzed 3–4% protein solutions. Therefore, until more data at lower concentrations is accumulated for a number of proteins, we suggest that a minimum absorbance value of 0.06 should be obtained for the amide I in order to ensure good repeatability of the nonlinear regression fit and ultimately the secondary structural model obtained for the protein.

## 4. Conclusions

FTIR analysis of lysozyme for its global secondary structural model in water ( $H_2O$ ) using nonlinear regression to obtain the component wavenumbers and areas is a precise methodology. The only caution to be observed for the methodology is the absorbance value obtained for the amide I envelope. We found the absorbance of the amide I should be at least 0.06 in order to obtain precise results using the current instrumentation we have (the Nicolet model 740 spectrophotometer). Newer generation spectrophotometers (for example the Nicolet 860) have better signal-to-noise ratios, and would probably be capable of precisely analyzing protein spectra with the amide I absorbance at values less than 0.06.

Our previous work showed the methodology used in this report to be accurate for the 14 proteins studied [1] when compared with the X-ray crystallographic results. Combined with the

Table 1  
Percent area values found

Wavenumber	Analysis day number					Average area $\pm$ cov	Average area $\pm$ cov
	1	2	3	4	5	Days 1-5	Days 1-4
1690	5.00	5.17	4.94	4.77	3.26	4.67 $\pm$ 17%	4.97 $\pm$ 3.4%
1684	11.67	11.49	11.96	12.01	11.73	11.77 $\pm$ 1.7%	11.78 $\pm$ 2.1%
1675	6.79	6.96	6.82	7.02	6.44	6.81 $\pm$ 2.9%	6.90 $\pm$ 1.6%
1667	11.61	11.16	11.47	11.50	11.70	11.49 $\pm$ 1.7%	11.44 $\pm$ 1.6%
1660	13.40	13.54	13.89	14.06	13.08	13.59 $\pm$ 2.9%	13.72 $\pm$ 2.3%
1653	13.34	13.75	12.99	13.25	15.67	13.80 $\pm$ 7.9%	13.33 $\pm$ 2.4%
1646	12.27	12.00	12.51	12.21	11.4	12.08 $\pm$ 3.1%	12.25 $\pm$ 1.7%
1638	12.52	12.46	12.05	11.86	15.27	12.83 $\pm$ 11%	12.22 $\pm$ 2.6%
1630	2.27	2.26	2.51	2.32	1.11	2.09 $\pm$ 33%	2.34 $\pm$ 5.1%
1625	11.11	11.21	10.86	11.00	10.34	10.90 $\pm$ 2.7%	11.05 $\pm$ 1.4%

precision analysis reported here, the methodology presented to analyze for the global secondary structure of proteins in H<sub>2</sub>O is both accurate and precise. We recommend that researchers carrying out secondary structural analysis of proteins using FTIR of the protein in H<sub>2</sub>O perform a validation analysis as reported here on a model protein (like lysozyme) before attempting to analyze proteins of unknown structure.

## References

- [1] T.F. Kumosinski and J.J. Unruh, *Talanta*, 43 (1996) 199-219.
- [2] S.N. Timasheff, in H. Peeters (Ed.), *Protides of the Biological Fluids*, 20th Colloquium, 1973, pp. 511-519.
- [3] J.L. Finney, *Phil. Trans. R. Soc. Lond. B.*, 278 (1977) 3.
- [4] W.K. Surewicz, H.H. Mantsch and D. Chapman, *Biochemistry*, 32 (1993) 389.
- [5] E. Goormaghtigh, V. Cabiaux J.M. Ruyschaert, in: H.J. Hilderson and G.B. Ralston (Eds.), *Subcellular Biochemistry*, Vol. 23: *Physicochemical Methods in the Study of Biomembranes*, 1994, pp. 329-450.
- [6] R.C. Davies, A. Neuberger and B.M. Wilson, *BBA*, 178 (1969) 294.
- [7] A. Dong, P. Huang and W.S. Caughey, *Biochemistry*, 29 (1990) 3303.
- [8] W. Damert, *Quantum Chem. Program Exchange Bull.*, 14 (4) (1994) 61.