FT-IR Studies of Protein Secondary Structure in Aqueous and Dried States

Ben Garland, Nicolet Spectroscopy Research Center, San Jose, CA, USA

INTRODUCTION

Proteins are polypeptides typically formed by more than 50 amino acids. The structural properties of a protein can be characterized by its primary, secondary, tertiary, and quaternary structures. The primary structure is determined by the sequence of amino acids that make up the protein. The secondary structure describes the orientation of the protein backbone and can be represented by structural elements such as alpha-helixes, beta-sheets, turns, and nonordered or irregular structures. The folding of the protein backbone upon itself to form a globule is called the tertiary structure. The quaternary structure is related to the interactions between certain protein subunits such as the four protein molecules called globins which form one unit of hemoglobin.

Proteins by nature are labile and can be denatured by different stresses in the aqueous, dried, or frozen states. Denaturation is the process in which the protein is rendered "dead" or inactive. This results from an unfolding of the protein's native secondary and tertiary structure. Unfolding can lead to subsequent aggregation. Aggregation can be detected by the resulting cloudy appearance of the protein solution.

The need for stabile protein formulations is growing rapidly, as the applications for recombinant and naturally derived proteins also continue to increase. The desired protein formulation is one that yields the maximum shelf lifetime by maintaining the protein in a native state while preserving the proteins activity.

Arakawa *et al*[1] have written an excellent review article discussing the factors which affect both the short and long-term protein stability. Generally, co-solutes like sugars have been used to stabilize the proteins. In this paper, they also discuss the thermodynamic relationships between the protein molecules and the co-solutes in solution and during freezing. In addition, they describe the "water substitute" mechanism that accounts for protection of dried proteins by sugars. An understanding of how to protect proteins during freeze-drying is important because many proteins are lyophilized for long-term storage.

To further understand the factors that influence protein stability, it is important to use instrumentation which can provide insight into the structure of the protein as well as monitoring any conformational changes induced by lyophilization.

It is very beneficial to determine if the protein in the dried formulation or the reconstituted protein has the same secondary conformational structure as that of the native protein. X-ray crystallography and multidimensional nuclear magnetic resonance (NMR) spectroscopy are both high resolution techniques for gaining insight into the three dimensional structure of proteins[2]. X-ray crystallography works well for those proteins in which a single crystal is available but cannot be used to accurately describe the protein conformation in a liquid environment, nor in the lyophilized solid. However, NMR can be used to study the secondary structure of the protein in a more biological like environment and in dried samples. The interpretation on the other hand becomes very complex and is limited to smaller proteins of less than 15-20 kDa.

Fourier transform infrared (FT-IR) spectroscopy is an established method for studying the secondary structure of polypeptides and proteins in both the solution and solid phases.[3] FT-IR is the only technique which can monitor the protein's structure in the dried state whereas, all other techniques provide results only on the rehydrated form of the protein. The secondary structure of a protein is reflected in the IR spectrum by the absorbances in the amide I region. The amide I band absorbs in the 1620 - 1690 cm^{-1} region and is primarily due to the stretching vibrations of the carbonyl groups. The absorbance bands around 1635 cm^{-1} are associated with the beta-structure whereas, the bands close to 1653 or 1646 cm^{-1} are associated with the helical portions and random portions of the protein, respectively.[4]

Recent studies have demonstrated that FT-IR analysis can be used as a predictive tool in developing effective formulations for stabilizing proteins during lyophilization.[5,6,7] This work has shown that a comparison of the second derivative spectra for both the lyophilized and liquid proteins can provide important information on the lyophilized protein activities.

This comparison is a quantitative means of obtaining the differences or similarities of the two proteins secondary structure in the two states. The correlation coefficient is calculated for the second derivative spectra over the amide I region of two protein samples. Identical spectra would provide a value of 1, while spectra of differing secondary structure would yield a value less than 1.

This work has also shown that the values critical for dried versus aqueous proteins correlates directly with the recovery of protein activity after rehydration. The calculation of the correlation coefficient provides an easy method for quantitatively comparing IR spectra of proteins in different environments. Thus, using FT-IR, one can easily study the effects that the protein formulation process, lyophilization, storage, and reconstitution have on the secondary structure of the protein.

This paper will discuss the instrumentation, accessories and software needed to obtain the protein second derivative spectra and correlation coefficients between spectra. Additionally, step by step procedures for sample preparation, IR data collection and processing will be presented. The protein chymotrypsinogen from Sigma was used in this paper to demonstrate the process required to perform this powerful yet simple IR experiment.

INSTRUMENTATION & ACCESSORIES

The instrument used was the Nicolet Magna-IR® 550 FT-IR spectrometer equipped with a KBr beamsplitter and DTGS room temperature detector. The bench contained the purge upgrade kit which includes a second purge line with

independent control of the purge gas flow into the back of the sample compartment.

Experimental results (from Carpenter's lab, see Acknowledgments) have indicated that a flow rate of 35 SCFH and 20 SCFH for the main bench and sample compartment respectively are sufficient to remove most of the water vapor from the system. Furthermore, the sample compartment was equipped with two KBr windows to help maintain purge conditions as well as minimizing the purge time between samples. Dry air or nitrogen purge is mandatory to prevent water vapor interference in the amide I region. A KBr pellet press and dye set is needed for preparing the KBr pellets of solid samples.

The cell that was used for the liquid analysis was similar to the Spectra Tech Econo liquid transmission cell and was equipped with a 10 micron polyethylene spacer and two CaF_2 windows. Cells like the Spectra-Tech demountable cell containing Leur-Lok Teflon® stoppered ports provide a more constant pathlength and are easier to fill than the previous cell type. The pathlength should be in the 6 to 10 micron range for any transmission cell used.

Totally absorbing water bands will be obtained with cell pathlengths greater than 10 microns. Additionally, high throughput single bounce ATR (attenuated total reflectance) accessories with ZnSe crystals can be used for highly concentrated protein solutions on the order of 50 mg/mL. The ATR can also be used to measure the lyopholized protein provided that the proper contact can be made between the protein and the ATR crystal.

ATR is a surface technique where the IR beam interacts only with the first couple of microns of the sample. Furthermore, the depth of penetration into the sample is wavelength dependant. This means that a greater depth of penetration is obtained at longer IR wavelengths (lower wavenumbers) while a shorter depth of penetration occurs at the shorter wavelengths (higher wavenumbers). Thus, the resulting ATR spectrum will exhibit different band intensities (lower absorbance values at higher wavenumbers) than the corresponding IR spectrum obtained by transmission. The advantages with the ATR analysis are the ease of use and ease of cleaning.

EXPERIMENTAL

Protein Preparation

Solids

Solid samples were and can be prepared by adding about 1 mg of the protein powder to about 300 mg of dry spectroscopic grade KBr powder. The KBr can be dried in an oven at about 110ºC for several hours. Prepare a homogeneous mixture of the protein with the KBr by mixing in a wigglebug or grinding with a mortar and pestle. Add the mixture to a KBr pellet dye set and press the pellet under vacuum for best results. Note: ideally, the pellet should be prepared in a dry box to reduce any effect that ambient water vapor might have on the hydration state of the protein.

Liquids

Liquid samples were and can be prepared by adding the protein sample to the appropriate buffer solution to obtain a concentration of about 20 mg/mL or greater. Results can be obtained on proteins of lower concentrations, with the lower limit at about 4 mg/mL. Use caution while reconstituting the protein since severe agitation will cause foaming and also cause the formation of protein aggregates.

The choice of an appropriate buffer is critical. The buffer must maintain an appropriate pH to prevent the protein from denaturing. Additionally, the buffer should not be IR active in the amide I region other than the water absorbance band centered at about 1644 cm^{-1}. Use a micro syringe or pipette to transfer a small amount (about 20 microliters) of the buffer solution to the sealed cell containing the 6 or 10 micron spacer. Collect the buffer spectrum which will be used to remove the buffer (water) spectral features from the protein spectrum by spectral subtraction. Use a vacuum to remove the buffer solution from the demountable type cell or dismantle and clean the Econo type cell. Then, inject the protein solution into the cell for subsequent IR analysis.

DATA COLLECTION

The data collection parameters should be set to collect 512 scans at 4 cm^{-1} (2 cm^{-1} data point spacing) resolution using Happ-Genzel apodization. Under these conditions, collect an open beam background which is a single beam measurement acquired without any sampling accessories in the beam path. This background spectrum will be used to collect subsequent absorbance sample spectra.

Note: it is critical to obtain a constant low level of water vapor in the background before using it to obtain absorbance spectra.

Next, collect the sample spectra using the same conditions. For the solid samples, collect the IR transmission spectrum of the protein KBr pellet. For the liquid samples, collect an IR transmission spectrum of the buffer solution and a second spectrum of the protein-buffer solution. The only compositional difference between the two solutions should be the addition of the protein to the sample solution.

The spectral results should resemble those shown in Figure 1. Here, a 50 mg/mL solution of the protein chymotrypsinogen (CTG) was prepared by dissolving it in 10 mM potassium phosphate buffer (pH=7.5). The top two spectra are of the buffer and 5% protein solutions, respectively, as acquired using the CaF_2 liquid transmission cell. The bottom spectrum was obtained by the transmission analysis of the protein KBr pellet.

Using a constant pathlength, the absorbance value for the band at 1640 cm^{-1} in the protein spectrum will be greater than the corresponding band in the buffer solution spectrum. This is caused by the

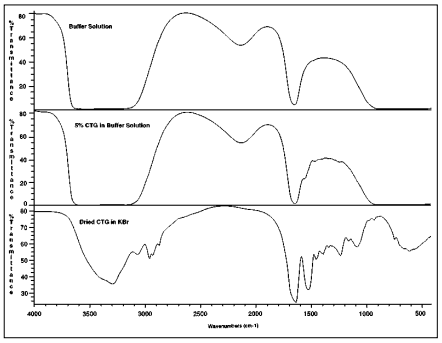

Figure 1. From top to bottom, the transmission spectra of the buffer solution (10 mM potassium phosphate buffer, pH = 7.5), protein solution (50 mg/mL of chymotrypsinogen) and the solid protein (KBr pellet). *Note: The data on aqueous vs. dried chymotrypsinogen came from Allison and Carpenter.*[8]

summation of the absorbances from both the protein amide I band and the water band which both absorb in this region.

Next, convert each spectrum to absorbance units for the subtraction of the aqueous buffer spectrum from the protein spectrum. The subtraction is accomplished by adjusting the factor to a value which removes all of the water from the protein spectrum. A good subtraction result is obtained when the broad water band at about 2100 cm^{-1} is zeroed as shown in Figure 2.

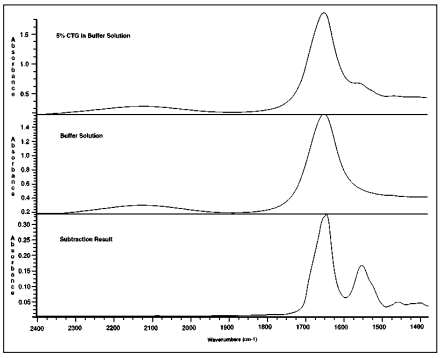

Figure 2.
The resulting spectrum (bottom) illustrates a good spectral subtraction of the buffer solution (middle) from the protein solution (top). The water band at about 2100 cm^{-1} was used for subtraction purposes. *Note: The data on aqueous vs. dried chymotrypsinogen came from Allison and Carpenter.*[8]

A water vapor spectrum will be needed for subsequent manipulations. Collect this spectrum using the same data collection parameters outlined above. In this case, the sample will be an open beam measurement which will be ratioed against the previously collected background spectrum. However, before collecting the sample spectrum, open and close the sample compartment door to allow water vapor to enter into the sample compartment. The resulting spectrum as shown in Figure 3 will be a 100% line containing both water vapor and carbon dioxide absorbance

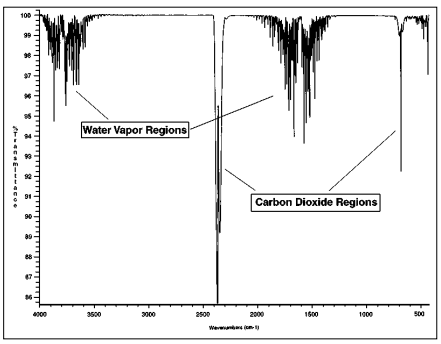

Figure 3.
A 100 % line water vapor spectrum. This spectrum, if necessary, is used to remove residual water vapor from the protein spectrum.

bands. This water vapor spectrum may be required to subtract interfering residual water vapor bands from the resulting protein spectrum shown in Figure 2. This subtraction can be accomplished before or after the second derivatives of each spectrum have been taken.

Take the second derivative of both the protein and water vapor spectra. Display both spectra in the same window from about 1800 - 1600 cm^{-1}. This region includes the Amide I band which provides valuable information on the secondary structure of the protein. If required, subtract the second derivative water vapor spectrum from the second derivative protein spectrum until the resultant spectrum is smooth in the 1800 cm^{-1} region or until the water vapor effects are minimized. The resulting spectrum shown in Figure 4, contains the information on the secondary structure of the protein for which the correlation coefficient can be calculated.

The correlation coefficient software interfaces with the Nicolet OMNIC® software

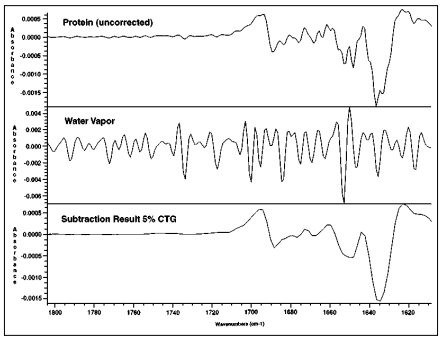

Figure 4.
The top spectrum is the second derivative spectrum of the protein containing a small amount of residual water vapor. The middle spectrum is the second derivative result of the water vapor spectrum. The bottom spectrum is the second derivative result of the protein with the water vapor removed.

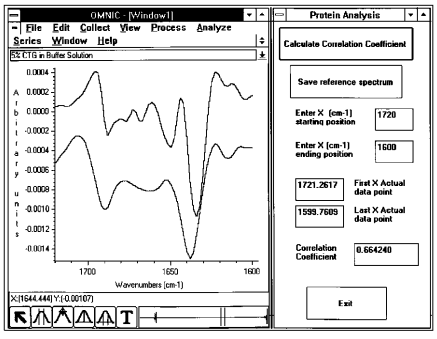

Figure 5.
This is the Windows™ based protein analysis program (right) which is coupled with the power of the Nicolet OMNIC (left) software package. A single click on the protein analysis icon will provide a user interface as shown above. *Note: The data on aqueous vs. dried chymotrypsinogen came from Allison and Carpenter.*[8] The correlation coefficient for these two spectra is 0.6642.

as shown in Figure 5. To run the entire analysis, one would simply select the protein icon. This will load both the OMNIC and protein analysis software programs. Data collection and manipulation is accomplished using the OMNIC software.

The protein analysis software easily allows the determination of the correlation coefficient between the previously saved reference spectrum and that of the current protein spectrum. The reference spectrum could be the second derivative spectrum of the native (aqueous) protein, whereas the sample could be the lyophilized (solid) protein The calculation is based on the default region of 1720 to 1600 cm^{-1}, which can be modified by typing in new limit values.

To calculate the correlation coefficient, simply select a reference spectrum from the OMNIC spectral display and click on the "save reference spectrum" button. Now, select the spectrum for comparison, and click on the "calculate correlation coefficient" button. The correlation coefficient result will be displayed in red. The proteins will have an increasingly similar secondary conformational structure as the correlation coefficient approaches unity.

CONCLUSION

Many protein formulation engineers and chemists can benefit from the powerful techniques of infrared spectroscopy. One can compare the solid state secondary structure with that of the native protein by calculating the correlation coefficient between the two IR spectra. This can also be done as a QC technique to monitor all protein formulations for structural changes.

ACKNOWLEDGMENTS

[1] Arakawa, Tsutomu; Prestrelski, Steven; Kenney, William; and Carpenter, John; *Advanced Drug Delivery Reviews*, **10**, 1-28 (1993).

[2] Surewicz, Witold K. ,Mantsch, Henry H., and Chapman, Dennis; *Biochemistry*, **32**, 389-394 (1993).

[3] Ismail, Ashraf; Mantsch, Henry; and Wong, Patrick; *Biochimica et Biophysica Acta*, 1121, **183-188** (1992).

[4] Byler, D.; Susi, Heino; *Biopolymers*, **25**, 469-487 (1986).

[5] Prestrelski, Steven; Tedeschi, Nicole; Arakawa, Tsutomu; and Carpenter, John; *Biophysical Journal*, **65**, 661-671 (1993).

[6] Carpenter, John; Prestrelski, Steven; Arakawa, Tsutomu; *Archives Of Biochemistry and Biophysics*, **303**, 456-464 (1993).

[7] Prestrelski, Steven; Arakawa, Tsutomu; Carpenter, John; *Archives Of Biochemistry and Biophysics*, **303**, 465-473 (1993).

[8] Allison, S. Dean and Carpenter, John; *Effects of Lyophilization in the Prensence of a Chaotroph on Aggregation of alpha-Chymotrypsinogen* (Manuscript in Preparation).

Additionally, I would like to acknowledge the following:

1. The assistance of John F. Carpenter, Ph.D.. (University of Colorado School of Pharmacy) and Steven J. Prestrelski, Ph.D. (Alza Corporation) for their critical comments on the paper.
2. John F. Carpenter for providing training in the described methods as well as the use of his lab and materials to obtain the spectra shown.