BIOPROCESS<u>TECHNICAL</u>

Fourier-Transform Midinfrared Spectroscopy for Analysis and Screening of Liquid Protein Formulations

Part 2: Detailed Analysis and Applications

Patrick Garidel and Heidrun Schott

n the first half of this two-part article (1), we introduced the physical principle behind infrared (IR) absorptions and experimental setups used for investigating protein samples. This second part concludes by focusing on the use of IR spectroscopic data in formulation stability studies and for characterizing protein secondary structures and related changes (e.g., resulting from protein–excipient interactions). We also examine IR's use in screening protein formulations.

BAND ASSIGNMENTS

As the structural repeat units of proteins, the peptide bond exhibits a number of IR-active amide bands (Table 1). Nine characteristic bands are termed amide A, B, and I–VII. Current understanding of IR spectra of proteins with regards to those

PRODUCT FOCUS: RECOMBINANT PROTEINS IN LIQUID FORMULATIONS

PROCESS FOCUS: PRECLINICAL TESTING AND PRODUCT FORMULATION

WHO SHOULD READ: R&D, ANALYTICAL, MANUFACTURING, PROCESS SCIENCE, AND FORMULATION STAFF

Keywords: Spectroscopy, infrared, analytical methods, formulations

LEVEL: ADVANCED

amide bands is largely based on normal coordinate analysis, which was pioneered by Miyazawa and associates in their milestone work on *N*methylacetamide (**2**, **3**) and subsequently extended to more complex systems.

The amide A (located at about 3400 cm⁻¹) and B (about 3090 cm⁻¹) bands originate from a Fermi resonance between the first overtone of amide II and the N-H stretching vibration (Figure 3 in Part 1 of this article). Amides I and II are the major bands in the protein IR spectrum. Amide I absorption originates from the C=0 stretching vibration (70-85%) of the amide group (coupled to in-phase bending of the N-H bond and stretching of the C-N bond), which gives rise to IR band(s) in the region between ~1600 and 1700 cm⁻¹. That vibrational mode is directly related to the protein "backbone" conformation. Amide II comes from the N-H bending (40-60%) and C-N stretching vibrations (18–40%) and is conformationally sensitive. Amides III and IV are very complex bands resulting from mixtures of several coordinate displacements. Out-ofplane motions are found in amides V, VL and VIL

Because of technical and theoretical limitations, only amide bands I–III are used for investigating protein secondary structure (4–7). Of all the



(WWW.PHOTOS.COM)

peptide amide modes, however, the most extensively used in such studies is amide I. Major factors responsible for conformational sensitivity include hydrogen bonding and couplings between transition dipoles (6, 7) (see Figure 2 in Part 1).

Transition dipole coupling that leads to splitting of the amide I mode. The magnitude of such splitting depends on the orientation of and distance between interacting dipoles. Thus it provides information about the geometrical arrangements of peptide groups in a polypeptide chain. Exact frequencies of amide I and II absorptions are influenced by the strength of hydrogen bonds involving amide C=O and N-H groups, as depicted by the equation in Figure 2 of Part 1 (1, 6, 7). Each element of secondary structural conformation (e.g., alpha helix or beta sheet) is associated with a characteristic

hydrogen bonding pattern between those groups. So each type of secondary structure gives rise to different frequencies at which amide bond vibrations occur, thus producing characteristic overall amide I and II vibrations (Figure 1). Such separation of subcomponent bands (Figure 2) for an overall amide I absorption forms the basis of protein structure analysis and quantification from vibrational spectroscopic data (8–10).

The relationship between amide I band position and type of secondary structure is best recognized by analyzing IR spectra of simple homopolypeptides that fold into welldefined and often homogeneous (purely alpha-helical or beta-sheet) structures. By contrast with such homopolypeptides, proteins usually fold into complex three-dimensional structures that include a variety of domains containing polypeptide segments folded into different types of secondary structures. Because each conformational entity contributes to the molecule's IR spectrum, observed amide I band contours are complex composites. They consist of many overlapping component bands that represent different structural elements, e.g., alpha helices, beta sheets, turns, and nonordered or irregular structures (Figure 2). A fundamental difficulty encountered in analyzing such composite band contours arises from the fact that the width of each contributing component band is usually greater than the separation between the maxima of adjacent peaks (Figure 2). So individual bands cannot be resolved and/or identified in the broad contours of experimentally measured spectra. Extraction of structural information encoded in those IR bands requires extensive mathematical analysis of experimental data.

In analyzing the amide I band, we must consider that some absorptions arising from buffer components may absorb between 1600 and 1700 cm⁻¹, thus perturbing analysis of the amide I protein absorptions. The most common interference is an absorption band of water in that region, but contributions of amino acid side chain absorptions must also be considered (11).

Other bands such as amide II have been used for the elucidation of structural information of proteins. Inclusion of that band with amide I has been reported by some authors using multivariate data analysis techniques to provide improved prediction accuracy (4, 12). Moreover, the amide II band has been used for monitoring hydrogen-deuterium exchange in proteins, providing information on subtle structural changes in protein secondary structure. Some authors have also used amide III band data to derive information with respect to protein secondary structure (7, 13, 14).

Figure 1 represents exemplarily different protein IR spectra obtained in transmission at protein concentrations of ~10 mg/mL in pure water at 25 °C. As can be deduced from the shape of the amide I absorption, secondary structural elements are different for these proteins. Hemoglobin shows a more or less symmetrical amide I absorption, whereas a broad amide I absorption is observed for alpha-lactalbumin and alcohol dehydrogenase. Such differences in overall amide I band shape come from different amounts of secondary structural elements in a protein structure.

ANALYSIS AND APPLICATIONS

Within the 1600–1700 cm⁻¹ IR spectral range, where the amide I absorption is detected, regions have been identified as being sensitive to particular secondary structural conformation. The "Band Assignments" box summarizes these structure sensitive regions within the amide I band. Hydrogen–deuterium exchange among amide hydrogen atoms causes a difference in band positions from those characteristically recorded in H₂O and D₂O.

Characterizing Protein Structure: Extraction of single-subcomponent bands is achieved using curve-fitting approaches to assign each subcomponent to a particular protein secondary structure. The principle of such procedures is to resolve an original protein structure into individual bands that fit its overall

Figure 1: FTIR spectra of various proteins in liquid formulation



Table 1: Band assignments of the main

 infrared active vibrations of biomolecules

⁶α-1-Proteinase inhibitor

³ a-Chymotrypsin

Frequency	
Range (cm ⁻¹)	Assignment
3490 and	Asymmetric and symmetric
3280	H–O–H stretching
3250-3300	Amide A (N–H stretch in
	resonance with amide II
	overtone)
3080	Amide B
3010	=C–H stretching of alkenes
2957	Asymmetric CH ₃ stretching
2920	Asymmetric CH ₂ stretching
2872	Symmetric CH ₃ stretching
2851	Symmetric CH ₂ stretching
1738	C=O stretch
1600–1700	Amide I (mainly C=O strech)
1645	H–O–H bending
1480–1575	Amide II (N–H bend in
	plane and C–N stretch)
1468	CH ₂ scissoring
~1395	C=O stretch of COO ⁻
1378	CH ₃ symmetric bend
1343	CH ₂ wagging
1230–1330	Amide III (N-H bend in
	plane and C-N stretch)
1240	Asymmetric PO ₂ ⁻ stretch
1170	Ester C–O asymmetric
	stretch
1080	Symmetric PO ₂ ⁻ stretch
1047	C-OP stretch
980	Choline asymmetric stretch
625–770	Amide IV (mainly O=C-N
	deformation)
720	CH ₂ rocking

spectrum. To estimate the range and positions of discrete subcomponent absorption bands, band-narrowing techniques are applied. Based on this Figure 2: An example of an amide I band with its underlying band subcomponents: alpha-helix (blue), beta-sheet (red), random coil (green)



procedure, secondary structure elements can be quantified.

One approach to extracting information on protein secondary structure from IR spectra is Fourier self-deconvolution, often referred as a "resolution enhancement" or "band narrowing" technique (1, 15-17). This procedure is based on an assumption that absorption bands are broadened (convoluted) in liquids by a certain function that causes a band overlap. A consequence is that the individual subcomponent bands cannot be distinguished in the overall amide I band. Fourier self-deconvolution decreases band widths, allowing separation of overlapping component bands that underlie a composite band's contour (16). This is also called deconvolution (17). The exact shape of the convolution function has not been determined. But it is assumed that Lorentzian and/or Gaussian functions are appropriate.

Two parameters are important for deconvolution: full width at halfheight (FWHH) and a resolution enhancement factor. In most cases, experienced operators determine these two parameters. Their selection determines the number and peak maximum of resulting subcomponent bands. It is a critical operation that determines the quality of results for a curve fitting procedure.

Overlapping band separation can also be increased by calculating the second derivative (Figure 3) of an absorption spectrum, either in the frequency domain or through mathematical manipulations in the Fourier domain (15, 18). Note that spectral derivation does not preserve the relative intensities of absorption bands. They depend on the width of each absorption in an original spectrum, so narrow bands will be enhanced at the expense of broader bands. A distinct advantage of the Fourier self-deconvolution method is that it introduces less distortion. Particularly, it does not affect the integrated intensities (areas) of individual component bands. So the effect is not to increase instrumental resolution, but rather the extent to which individual component bands can be separated.

The degree of band narrowing (described by the resolution enhancement factor in Fourier deconvolution and the degree of derivation in derivative spectroscopy) is limited by a spectrum's signal/noise ratio. So analysts should avoid "overdeconvolution" or using higher derivatives, because noise will also be amplified and can be easily misinterpreted as a real band (10).

Another challenge can arise from the presence of atmospheric water vapor, which gives rise to narrow absorption bands especially in the region overlapping the main protein bands. Although the overlapping bands are often very weak in an original spectrum, their relative sharpness makes them disproportionally amplified upon either Fourier deconvolution or derivation. They may appear in a resolutionenhanced spectrum as artifacts that are indistinguishable from the real components of a protein band. This problem can be elegantly eliminated by technical solutions such as purging spectrometers with dry air. If necessary, residual water vapor absorption may be compensated for by spectral subtraction.

Figure 2 shows determination of secondary structural elements for a chemokine at a 1 mg/mL concentration in PBS buffer at pH 7.4 (H_2O , 25 °C). A Gaussian, Lorentzian, or mixed Gaussian–Lorentzian shaped curve is determined that best fits the original protein spectra (the black curve represents the original spectrum, and the magenta curve represents overall

fit). To get initial parameters (band height, width, position, and baseline) for starting a fitting procedure for the individual subcomponent bands, Fourier self-deconvolution was used here. The best fit was obtained by a root mean-square analysis determining the optimal set of curve-fitting parameters, then the corresponding subcomponent band was determined. The area of each individual band is used to calculate its relative contribution to a particular protein secondary structure in relation to the overall area of the original spectrum.

In the example, the overall amide I band was fitted assuming just three secondary structural elements: alphahelical, beta-sheet, and random coil structures. Analysis revealed that the protein structure is about 45% alphahelical, 39% beta-sheet, and 16% random coil structures. Those results are in accordance with X-ray data. Our procedure assumes that the carbonyl molar absorption coefficient is equal for each individual structure.

Other possibilities for the determination of secondary protein structure are based on pattern

Figure 3: α-Lactalbumin in the respective presence and absence of magnesium ions; (TOP) IR spectra, (BOTTOM) second derivative of the amide I region



recognition. These methods use calibration matrices of IR spectra for proteins of known secondary structure (usually determined by X-ray crystallography). One matrix consists of target secondary structure fractions for proteins in a reference set; the other consists of their absorption data. The combined matrices serve as a calibration set. Based on that "training" set, a multivariate regression model is derived for predicting secondary structure elements of a "new" protein. The most widely used regression methods include factor analysis, principal component analysis, and singular value decomposition (17-22). Partial least-square methods and the methodology of factor analysis and multiple linear regressions are both routinely used, being implemented in commercially available software. In recent years, neural network analysis and two-dimensional correlation spectroscopy have become more important (23-26).

Testing Formulations: A number of excipients are known to bind to proteins or affect their stability. Protein–excipient interactions can easily be investigated using IR spectroscopy. For example, divalent magnesium cations interact with alpha-lactalbumin (Figure 3). Figure 3 TOP shows the original spectra (10 mg/mL, phosphate buffer, pH 6.2, 25 °C), and Figure 3 воттом shows second-derivative spectra. The presence of magnesium cations induces a distortion of the amide I band, as illustrated by the appearance of a "shoulder" at about ~1630 cm⁻¹. It is clearer in the second derivative (Figure 3 воттом). IR spectroscopic analysis of the amide I absorption revealed magnesium binding to the protein and inducing changes to its secondary structure.

In a number of cases during biopharmaceutical development, analysts may be interested only in recording differences between various protein formulations. The idea of such analysis is to compare an initial state of the protein product with its structures in a selected formulation e.g., after being stressed by temperature or pH. Various approaches have been described, with "visual" comparison of IR spectra in the 1700–1200 cm⁻¹ spectral region (amide I, II, and III absorptions) often used to identify spectral deviations. However, it is still unclear what constitutes "large" or "small" differences. In most approaches, correlation coefficients are calculated, which express the common features and similarities of the two compared spectra (27-29). A correlation coefficient for identical spectra can be set to 1, with the correlation coefficient at 0, when both spectra have nothing in common. The area overlap method implemented by several different groups uses the integrated difference between two area-normalized spectra (Figure 4). Each approach uses both raw and resolution-enhanced spectra.



Another useful method for detection of small changes in protein structure is difference spectroscopy, which involves subtracting a protein absorbance spectrum in one state from that of the same protein in another state. For example, the FTIR spectrum of a sample is obtained before and after being triggered by a particular effect that induces two different states of the protein (e.g., pH change, light, heat, or addition of an interaction partner). This approach reduces the complexity of interpreting conformational protein changes induced by the trigger, and the difference spectrum reflects only those groups that undergo a specific change

Figure 4: Example of an area overlap of two raw spectra. The area overlap of the two spectra (90 %) is indicated in yellow.



in their structure or environment (Figure 5).

Figure 5A shows the results from a FTIR-HATR experiment representing thermal-induced denaturation of an antibody at a concentration of 50 mg/mL. Up to a temperature of 60 °C, amide I changes are marginal (data not shown). Above 60 °C, however, clear changes are seen: The overall band intensity increases considerably. More important are changes in the band shape, indicating formation of intermolecular beta-sheet structures. Using the spectrum taken at 25 °C as a reference, derivative spectra could be calculated. These are plotted in Figure 5B, showing changes induced by thermal stress. The area of those changes can be calculated and plotted as a function of temperature.

Such protocols are used for fast screening of protein formulations as a function of temperature. The experiment is especially suited for analyzing highly concentrated protein formulations. Figure 6 shows denaturation curves for a 60-mg/mL protein solution at three different pH values. Denaturation temperature can be determined from the curve's sigmoidal shape. In the example, a formulation at pH 4.0 considerably increases thermally induced stability of the protein. Data are somewhat different when you compare the same formulations at much lower protein concentrations. IR spectroscopy thus allows direct investigation of highly concentrated protein formulations in their original protein concentration. Using highly sensitive calorimetric

Figure 5: (LEFT) Structural changes in the spectral region between 1600 and 1720 cm⁻¹ representing the amide I absorption as a function of temperature; (RIGHT) difference spectroscopy (temperature dependence of the amide I absorption)



Figure 6: Formulation screening of a protein solution at 60 mg/mL in water at different pH values (temperature dependence of the amide I)



methods, protein concentration is usually limited to 0.1-10 mg/mL (depending on the protein involved).

IR spectroscopy is very helpful for a number of other applications of elucidating structural protein changes in formulation. Information concerning formation of protein aggregates can also be derived from FTIR measurements (30-33). Appearance of strong absorption bands below 1620 cm⁻¹ often correlates with aggregation - usually associated with formation of strong beta-sheet structures. However, absence of such features in the amide I region does not indicate an absence of protein aggregates (34).

A FORMULATOR'S COOL TOOL

Fourier-transform IR spectroscopy is a powerful technique for structural and conformational characterization of proteins because of its ability to test samples under different physical conditions: aqueous and nonaqueous solutions as well as dry samples.

BAND ASSIGNMENTS

Band assignments of the subcomponent bands of the amide I absorption (data for a protein in H_2O : 1620-1640 cm⁻¹: Beta sheets 1640–1650 cm⁻¹: Nonordered structures 1650–1658 cm⁻¹: Alpha helices 1660-1680 cm⁻¹: Loops 1670-1695 cm⁻¹: Beta sheets

Measurements can be done for liquid formulations containing 1 to at least 100 mg/mL protein.

FTIR is well suited for studying protein stability in the development of protein formulations. It finds application for

• fast determination of protein secondary structure in solution or dry

biopharmaceutical quality control

 biomolecular interactions (protein-ligand binding) monitoring

• protein determination and concentration measurements

• study of aggregation and fibrillation processes

 detection of conformational changes (e.g., due to altered pH or ionic strength)

• investigation of temperature and buffer influence

 formulation screening for highly concentrated protein formulations.

ACKNOWLEDGMENTS

The authors are grateful to the people in their group, to S. Bassarab for support, and to W. Kliche for critical reading of this manuscript and useful suggestions.

REFERENCES

1 Garidel P, Schott H. Fourier-Transform Midinfrared Spectroscopy for Analysis and Screening of Liquid Protein Formulations: Part 1, Understanding Infrared Spectroscopy of Proteins. BioProcess International 4(5) 2006: 40 - 46

2 Miyazawa T, Shimanouchi T, Mizushima J. Normal Vibrations of N-Methylacetamide. J. Chem. Phys. 29 1958: 611-616.

3 Miyazawa T. Poly-α-Amino Acids. Fasman GD, editor. Marcel Dekker: New York, NY, 1967; 69-103.

4 Dousseau F, Pezolet M. Determination of the Secondary Structure Content of Proteins in Aqueous Solutions from Their Amide I and Amide II Infrared Bands: Comparison Between Classical and Partial Least-Squares Methods. Biochem. 29, 1990: 8771-8779.

5 Barth A, Zscherp C. What Vibrations Tell Us About Proteins? Quart. Rev. Biophys. 35, 2002: 369-430.

6 Bandekar J. Amide Modes and Protein Conformation. Biochim. Biophys. Acta 1120, 1992: 123-143.

7 Krimm S, Bandekar J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. Adv. Protein Chem. 38, 1986: 181-364.

8 Byler DM, Susi H. Examination of the Secondary Structure of Proteins By Deconvolved FTIR Spectra. Biopolymers 25, 1986: 469-487.

9 Oberg KA, Ruysschaert JM, Goormaghtigh E. The Optimisation of Protein Secondary Structure Determination with Infrared and Circular Dichroism Spectra. Eur. J. Biochem. 271, 2004: 2937-2948.

10 Haris PI, Chapman D. Does Fourier Transform Infared Spectroscopy Provide Useful Information on Protein Structures? TIBS 17, 1992: 328-333.

11 Venyaminov SY, Kalnin NN. Quantitative IR Spectrophotometry of Peptide Compounds in Water (H2O) Solutions: I -Spectral Parameters of Amino Acid Residue Absorption Bands. Biopolymers 30, 1990: 1243-1257.

12 Rahmelow K, Hübner W. Secondary Structure Determination of Proteins in Aqueous Solution By Infrared Spectroscopy: A Comparison of Multivariate Data Analysis Methods. Anal. Biochem. 241, 1996: 5-13.

13 Schweitzer-Stenner R, et al. Structure Analysis of Dipeptides in Water By Exploring and Utilising the Structural Sensitivity of Amide III By Polarised Visible Raman, FTIR-Spectroscopy, and DFT Based Normal Coordinate Analysis. J. Phys. Chem. B 106, 2002: 4294-4304.

14 Fu FN, et al. Secondary Structure Estimation of Proteins Using the Amide III Region of Fourier Transform Infrared Spectroscopy: Application to Analyse Calcium-Binding Structural Changes in Calsequestrin. Appl. Spectrosc. 48, 1994: 1482-1441.

15 Cameron DG, Moffatt DJ. Deconvolution, Derivation, and Smoothing of Spectra Using Fourier Transforms. *JTEVA* 12, 1984: 78–85.

16 Kauppinen JK, et al. Fourier Self-Deconvolution: A Method for Resolving Intrinsically Overlapped Bands. *Appl. Spectrosc.* 35, 1986: 271–276.

17 Powell JR, Wasacz FM, Jakobsen RJ. An Algorithm for the Reproducible Spectral Subtraction of Water from the FTIR Spectra of Proteins in Dilute Solutions and Adsorbed Monolayers. *Appl. Spectrosc.* **40**, 1986: 339–344.

18 Pistorius AMA, DeGrip WJ. Deconvolution As a Tool to Remove Fringes from an FT-IR Spectrum. *Vibr. Spec.* 36, 2004: 89–95.

19 Savitzky A, Golay MJE. Smoothing and Differentiation of Data By Simplified Least Square Procedures. *Anal. Chem.* 36, 1964: 1627–1639.

20 Lee DC, et al. Determination of Protein Secondary Structure Using Factor Analysis of Infrared Spectra. *Biochem.* 29, 1990: 9185–9193.

21 Wallace BA, Cascio M, Mielke DL. Evaluation of Methods for the Prediction of Membrane Protein Secondary Structures. *Proc. Natl. Acad. Sci. USA* 83, 1986: 9423–9427.

22 Oberg KA, Ruysschaert JM, Goormaghtigh E. Rationally Selected Basis Proteins: A New Approach to Selecting Proteins for Spectroscopic Secondary Structure Analysis. *Protein Sci.* 12, 2003: 2015–2031.

23 Hering JA, Innocent PR, Haris PI. Beyond Average Protein Secondary Structure Content Prediction Using FTIR Spectroscopy. *Appl. Bioinf.* 3, 2004: 9–20.

24 Hering JA, Innocent PR, Haris PI. Automatic Amide I Frequency Selection for Rapid Quantification of Protein Secondary Structure from Fourier Transform Infrared Spectra of Proteins. *Proteomics* **2**, 2002: 839– 849.

25 Hering JA, Innocent PR, Haris PI. Neuro-Fuzzy Structural Classification of Proteins for Improved Protein Secondary Structure Prediction. *Proteomics* 3, 2003: 1464– 1475.

26 Noda I. Advances in Two-Dimensional Correlation Spectroscopy. *Vibr. Spec.* 36, 2004: 143–165.

27 Prestrelski SJ, et al. Dehydration-Induced Conformational Transitions in Proteins and Their Inhibition By Stabilizers. *Biophys. J.* 65, 1993: 661–671.

28 Kendrick BS, et al. Quantification of the Area of Overlap Between Second-Derivative Amide I Infrared Spectra to Determine the Structural Similarity of a Protein in Different States. *J. Pharm. Sci.* 85, 1996: 155–158.

29 van de Weert M, et al. Fourier Transform Infrared Spectrometric Analysis of Protein Conformation: Effect of Sampling Method and Stress Factors. *Anal. Biochem.* 297, 2001: 160–169. **30** Schüle S, et al. Determination of the Secondary Protein Structure of IgG1 Antibody Formulations for Spray Drying. *Resp. Drug Deliv.* **2**, 2004: 377–380.

31 Schüle S, et al. Spray Drying of IgG1 Antibody Formulations: A CD and FTIR Investigation with Respect to Protein Secondary Structure. *Resp. Drug Del.* **1**, 2005: 185–188.

32 Lenk TJ, et al. IR Spectral Changes of Bovine Serum Albumin Upon Surface Adsorption. *J. Biomed. Materials Res.* 23, 1989: 549–569.

33 Kendrick BS, et al. Aggregation of Recombinant Interferon Gamma: Kinetics and Structural Transitions. *J. Pharm. Res.* 87, 1998: 1069–1076.

34 Schüle S, et al. Conformational Analysis of Protein Secondary Structure During Spray-Drying of Antibody/Mannitol Formulations. *Eur. J. Pharm. Biopharm.* 2006; submitted.

Corresponding author Patrick Garidel

(Dr. rer. nat.) is head of pharmaceutical basic development, and **Heidrun Schott** (Dipl.-Ing. FH) is responsible for developing infrared techniques in the pharmaceutical development group of the Department of Process Science/Biopharmaceuticals at Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach an der Riss, Germany; 49-7351-545556, fax 49-7351-835556; patrick. garidel@bc.boehringer-ingelheim.com, www.boehringer-ingelheim.com.