

# Quantitative Studies of Protein Structure by FT-IR Spectral Deconvolution and Curve Fitting

D. Michael Byler, Janine N. Brouillette, and Heino Susi U.S. Department of Agriculture\*

The peptide carbonyl stretching band (amide I band) of the infrared spectrum of a protein provides a sensitive probe of its secondary structure. If the spectrum has a sufficiently high signal-to-noise ratio, Fourier deconvolution can partially resolve the band into several components. These components can be associated with distinct segments of the peptide backbone that exhibit different conformations such as extended chains, helices, and various turns and bends. Gauss-Newton iterative curve fitting of the deconvoluted band affords a means of evaluating the areas of each component. Dividing the sum of the areas of all components associated with a given conformation by the total amide I band area gives a number that indicates what fraction of the protein has this particular conformation. The results obtained in this manner for helix content and for extended chain content of proteins in aqueous solution are in good agreement with values obtained from single-crystal x-ray data.

The peptide linkages of proteins exhibit several vibrational bands between 300 and 1700  $\text{cm}^{-1}$  (usually denoted as amide I to amide VII) that are characteristic of secondary *trans* amides. Of these vibrational bands, the amide I band — resulting primarily from the stretching vibration of the peptide carbonyl group — is the most sensitive to changes in the secondary structure or conformation of proteins (1). For many years, only qualitative structural information could be obtained from such spectra (1-5). More recently, quantitative studies have been attempted by curve fitting the spectral data with the help of digital computers (6). Modern instrumentation, particularly Fourier

\*Reference to a brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over other instruments of a similar nature that are not named.

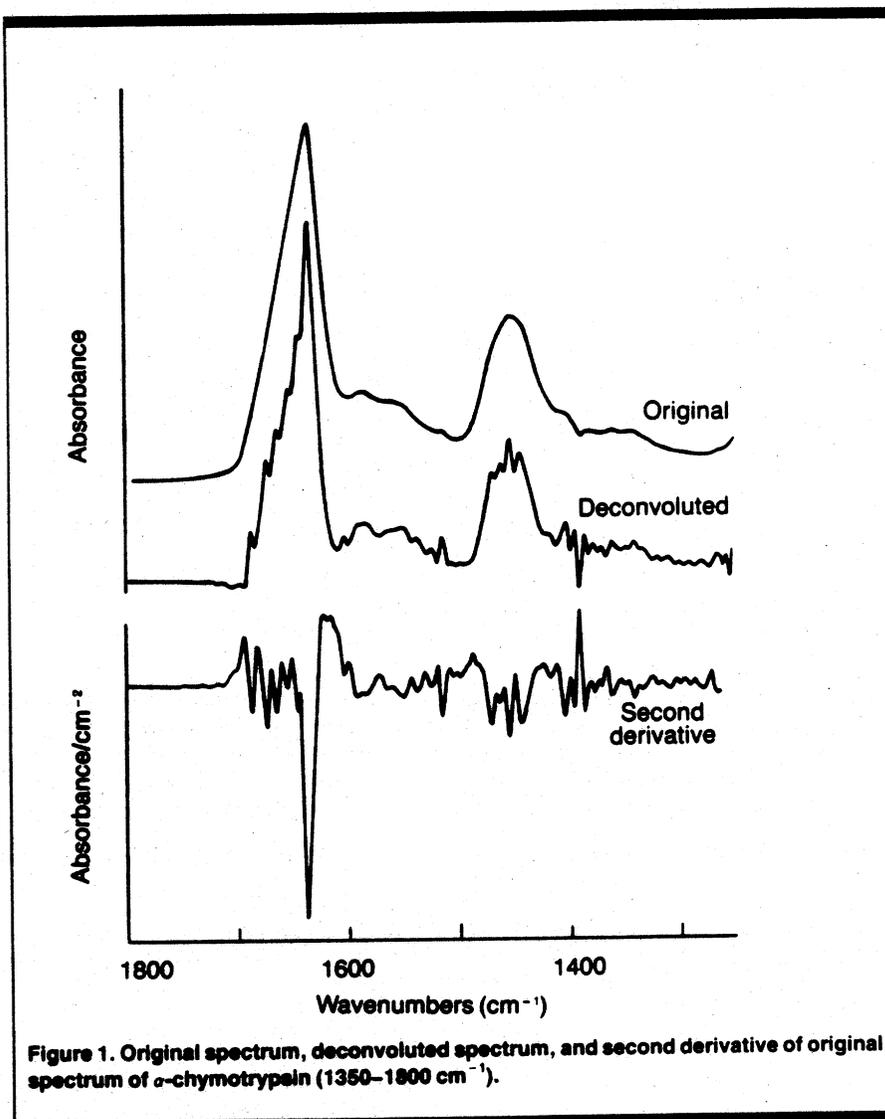


Figure 1. Original spectrum, deconvoluted spectrum, and second derivative of original spectrum of  $\alpha$ -chymotrypsin (1350–1800  $\text{cm}^{-1}$ ).

transform interferometers, now can produce data with such low noise levels that one can reliably use resolution-enhancement techniques — such as differentiation and Fourier deconvolution — to obtain spectral detail that may not have been easy to extract from the

original spectral data otherwise (7,8). In particular, when these methods are applied to the broad amide I band of globular proteins in deuterium oxide solution, a number of component features are disclosed (Figure 1). The components result from different conforma-

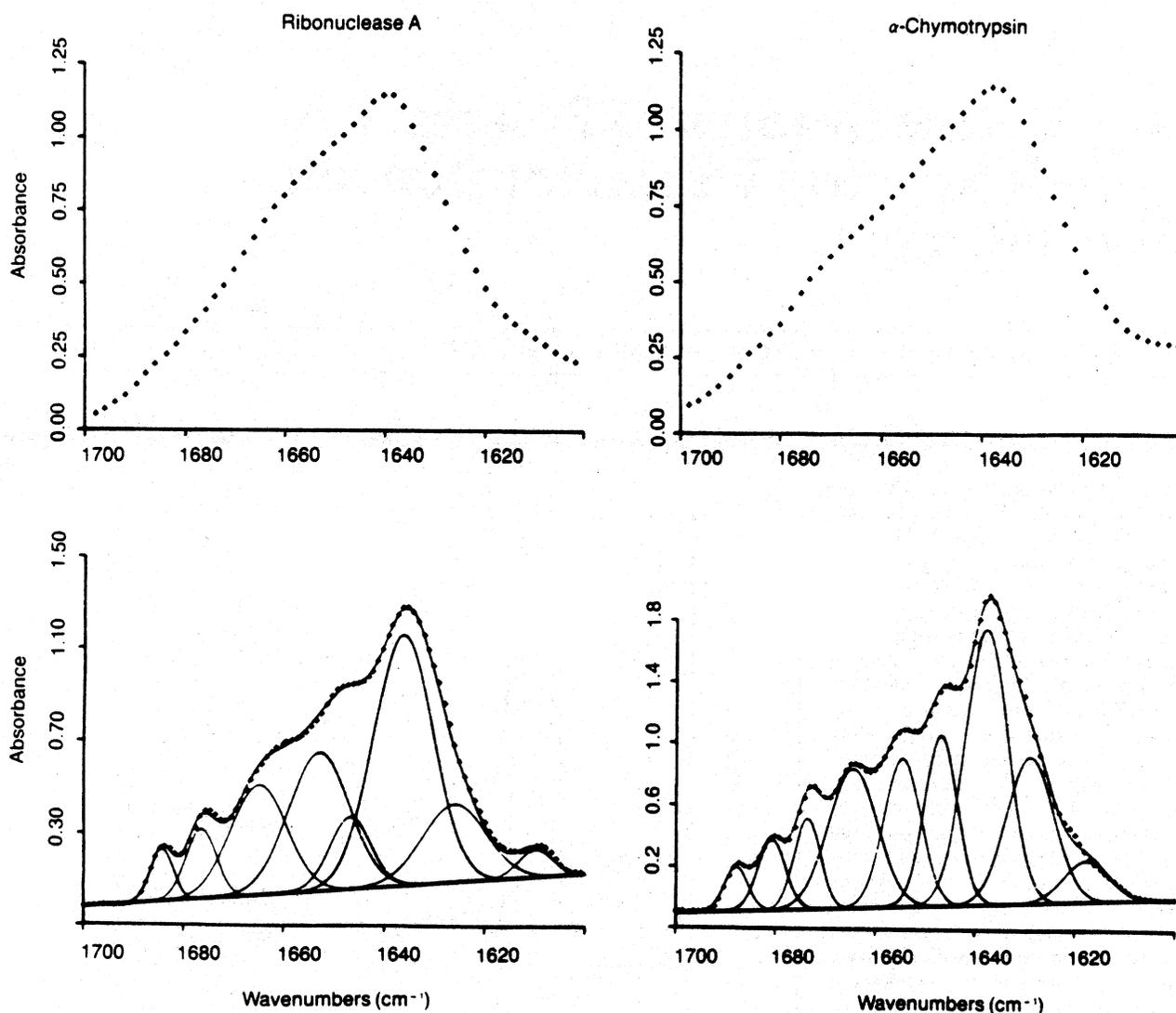


Figure 2. Original digitized spectra (upper curves) and deconvoluted spectra and fitted amide I components (lower curves) for ribonuclease A and  $\alpha$ -chymotrypsin. Absorbance is plotted versus wavenumber. In the top curves, every second point of the digitized spectrum is shown as a cross. In the lower curves, crosses represent the experimental deconvoluted spectrum; the solid lines give the resolved components and the calculated sum of the resolved components. (Because the deviation between the calculated curve and the experimental data is so small, the former is virtually superimposed on the experimental spectrum.)

tional substructures that make up the backbone of the proteins. More-detailed quantitative information concerning the protein secondary structure can be obtained by combining the results from Fourier deconvolution of the amide I band with an iterative curve-fitting procedure.

#### EXPERIMENTAL

Proteins were obtained from Sigma Chemical Company (St. Louis, Missouri). Sufficient protein was dissolved in 0.01 M NaCl/D<sub>2</sub>O to produce protein concentrations of about 5% w/v. The solutions were allowed to equilibrate at ambient temperature for about 24 h

to ensure maximum isotopic exchange. The degree of isotopic exchange was judged by the disappearance of the amide II band around 1550 cm<sup>-1</sup>. Upon deuteration, the contribution of the N-H bending vibration to the amide II mode causes the band to shift about 100 cm<sup>-1</sup>. The use of D<sub>2</sub>O solutions for quantitative studies is necessary because H<sub>2</sub>O has a strong absorption near 1640 cm<sup>-1</sup> (the HOH bending mode) that overlaps the amide I mode of the protein (1620–1700 cm<sup>-1</sup>). Accurate spectral subtraction of the intense solvent band from the weaker protein absorption is therefore virtually impossible (1). Fortunately, O- and N-deuteration of a protein produces rather small isotopic frequency shifts

( $\leq 10$  cm<sup>-1</sup>) of the amide I protein band while shifting the bending vibration of water nearly 400 cm<sup>-1</sup>. Therefore, all protein spectra for this study were obtained using D<sub>2</sub>O solutions.

FT-IR spectra from 1300 to 1800 cm<sup>-1</sup> were measured with a model 7199 spectrometer (Nicolet Analytical Instruments, Madison, Wisconsin). All spectra were obtained from 4000 coadded interferograms to secure signal-to-noise ratios of better than 1000:1. The signal-averaged interferogram was zero-filled once and apodized with the Happ-Genzel function prior to phase correction and

**Table I. Amide I band components observed in the deconvoluted FT-IR spectrum of  $\alpha$ -chymotrypsin (1620–1700  $\text{cm}^{-1}$ ).**

Frequency ( $\text{cm}^{-1}$ )	Assignment	Fraction of amide I band area (%)
1688	Turns	2.3
1680	Turns	5.7
1673	Extended chain	5.4
1665	Turns	15.5
1655	$\alpha$ -Helix	11.7
1647	Unordered	15.3
1638	Extended chain	25.7
1630	Extended chain	18.4

Fourier transformation (9–11). Nominal instrument resolution was  $2 \text{ cm}^{-1}$ , and the cell path length was 0.075 mm. Deconvolution was performed with a Nicolet program that is based on the algorithm of Kauppinen and colleagues (9–11). In addition to selecting the region of the spectrum on which to operate, the experimenter must specify the values of two parameters: VF0, the width at half height of the line-shape function with which the band is to be deconvoluted; and VF1, a resolution-enhancement factor. In the referenced work (9–11), VF0 and VF1 correspond to  $2\sigma$  and K, respectively.

For protein solution spectra deconvoluted over the range of  $1350\text{--}1800 \text{ cm}^{-1}$ , it was determined that setting VF0 at  $13 \text{ cm}^{-1}$  and VF1 at 2.4 generally gave optimal resolution enhancement without producing evidence of excessive deconvolution, such as side lobes or other artifacts (8,12). Under such conditions, the areas of the component bands are unchanged by deconvolution because the height of each peak increases as the width decreases (9–11). It should be emphasized that excessive deconvolution can distort the areas of overlapping bands and must be carefully avoided if quantitative information is to be obtained from deconvoluted spectra.

In preparation for curve fitting, band center frequencies were confirmed by second derivative spectroscopy (7). Figure 1 shows that the second derivative spectrum and the deconvoluted spectrum of a protein yield nearly identical frequency values for all resolution-enhanced bands. These frequencies cannot be determined from the original spectrum.

Curve fitting was accomplished with the program ABCUS, written by W.C. Damert of ERRC, U.S. Department of Agriculture. Gaussian bandshapes were assumed for the deconvoluted components. Although this approximation has not yet been confirmed by

theory, it does lead empirically to quite satisfactory results. Iterative curve fitting, however, does not always lead to a unique solution. Nevertheless, "false" solutions usually can be easily detected and rejected because they often involve unreasonably large values for one or more bandwidths and a relatively large root-mean-square (RMS) error.

## RESULTS AND DISCUSSION

Figure 2 shows the original spectra, the deconvoluted spectra, and the fitted components for the amide I bands of ribonuclease A and  $\alpha$ -chymotrypsin, two globular proteins that are rich in  $\beta$ -structure. In both cases, the RMS deviation between the calculated and the observed deconvoluted spectrum is less than 1% of the maximum absorbance value. Table I presents the observed frequencies, vibrational assignments, and relative areas for each amide I band component of  $\alpha$ -chymotrypsin. The assignments are based on previous work in the authors' laboratory (1,2,6–8) and on normal coordinate calculations by Krimm and co-workers (3,13). Based on the frequency distribution of 138 observed amide I band components for 19 globular proteins with widely varying structures (12,15), a recurring pattern of bands at a limited number of distinct frequencies becomes apparent. As shown in Table II by the mean frequency values and absolute ranges of each, these 11 characteristic amide I components show relatively little overlap with each other. The table also presents proposed assignments for each characteristic frequency (12). Table III gives the percentage of  $\alpha$ -helix content and extended-chain content for six typical proteins. These percentages were calculated from FT-IR data that were similar to the data listed in Table I for  $\alpha$ -chymotrypsin. Corresponding values for the secondary structure of these proteins — taken from Levitt and Greer's comprehensive examination of protein x-ray data (16) — are provided for comparison. This study provides an extremely thorough and consistent interpretation of protein crys-

**Table II. Characteristic frequencies ( $\text{cm}^{-1}$ ) and assignments of amide I components\* for 19 globular proteins.**

Mean frequency ( $\text{cm}^{-1}$ )	Assignment
$1624 \pm 4^\dagger$	Extended chain
$1631 \pm 3$	Extended chain
$1637 \pm 3$	Extended chain
$1645 \pm 4$	Unordered
$1653 \pm 4$	$\alpha$ -Helix
$1663 \pm 4$	Turns
$1671 \pm 3$	Turns
$1675 \pm 5$	Extended chain
$1683 \pm 2$	Turns
$1689 \pm 2$	Turns
$1694 \pm 2$	Turns

\*See references 12 and 15.

†Maximum range of the observed frequency for each component.

tallographic data and covers a total of 62 proteins. It is not possible to provide similar comparisons for bends and turns, because these protein substructures do not have the same type of periodic, repeating structures as those found in  $\alpha$ -helices and  $\beta$ -sheets. In addition, bends and turns usually consist of only four or less adjacent residues. For these reasons, it is still not possible to determine whether a specific amino acid residue in a given protein belongs to a turn rather than an end of a helix or a  $\beta$ -strand.

As seen in Table III, the agreement between the FT-IR results and values derived from x-ray crystallography is quite good. It should be noted that results from solution data in Table III are compared with those from solid state values. At present, this is unavoidable because no uniform, comprehensive set of secondary structure data for proteins in aqueous solutions has been presented.

One cannot emphasize too strongly that values for "percent helix" or "percent extended structure" are always somewhat subjective, even when the calculations are based on accurate bond lengths and angles. Ambiguity arises from the uncertainty in choosing the exact point along the peptide chain where one segment of a secondary structure begins and another ends. This choice depends not only on the manner in which an ideal helix or sheet has been defined by the various investigators (12,16) but also on just how regular any of these regions of protein substructure really are. With these uncertainties in mind,

**Table III. Percent helix and extended chain calculated using FT-IR and x-ray analysis for six typical proteins.**

Protein	Helix content (%)		Extended chain (%)	
	FT-IR	X-ray*	FT-IR	X-ray*
Carboxypeptidase	40	39	33	30
$\alpha$ -Chymotrypsin	12	10	50	49
Concanavalin A	4	2	60	60
Lysozyme	41	45	21	19
Papain	27	29	32	29
Ribonuclease A	21	22	50	46

\* Obtained from reference 16.

the agreement between the reported x-ray values and the FT-IR values is particularly encouraging.

Despite these caveats and limitations, analysis of protein secondary structure in aqueous solution by curve fitting of deconvoluted FT-IR spectra offers great promise for a large number of future studies. A previous investigation already has shown that one can observe the qualitative changes in protein conformation that occur when the environment of the protein is altered by the addition of polar organic solvents, such as alcohols (17). In addition, work currently in progress in the authors' laboratory indicates that structural changes resulting from the complexation of metal ions by specific side groups can also be detected. Finally, alterations in the folding of the peptide backbone as a result of variations in temperature or pH are clearly discernible. In each case, the combined use of deconvolution and curve fitting can provide quantitative estimates of the structural changes involved.

#### REFERENCES

- (1) H. Susi, *Meth. Enzymol.* **26**, 455-472 (1972).
- (2) S.N. Timasheff, H. Susi, and L. Stevens, *J. Biol. Chem.* **242**, 5467-5473 (1967).
- (3) S. Krimm, *J. Mol. Biol.* **4**, 528-540 (1962).
- (4) T. Miyazawa, *J. Chem. Phys.* **32**, 1647-1652 (1960).
- (5) A. Elliott and E.J. Ambrose, *Nature* **165**, 921-922 (1950).
- (6) M. Rüegg, V. Metzger, and H. Susi, *Biopolymers* **14**, 1465-1471 (1975).
- (7) H. Susi and D.M. Byler, *Biochem. Biophys. Res. Comm.* **115**, 391-397 (1983).
- (8) W.-J. Yang, P.R. Griffiths, D.M. Byler, and H. Susi, *Appl. Spectrosc.* **39**, 282-287 (1985).
- (9) J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, and D.G. Cameron, *Appl. Spectrosc.* **35**, 271-276 (1981).
- (10) J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, and D.G. Cameron, *Anal. Chem.* **53**, 1454-1457 (1981).
- (11) J.K. Kauppinen, *Spectrometric Tech.* **3**, 199-232 (1983).
- (12) D.M. Byler and H. Susi, *Biopolymers* **25**, in press (1986).
- (13) S. Krimm and J. Bandekar, *Biopolymers* **19**, 1-29 (1980).
- (14) A.M. Dwivedi and S. Krimm, *J. Phys. Chem.* **88**, 620-627 (1984).
- (15) D.M. Byler and H. Susi, unpublished spectra.
- (16) M. Levitt and J. Greer, *J. Mol. Biol.* **114**, 181-239 (1977).
- (17) J.M. Purcell and H. Susi, *J. Biochem. Biophys. Meth.* **9**, 193-199 (1984).

*D. Michael Byler is a research chemist in the plant and soil biophysics research group at the U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center (USDA-ERRC), 600 East Mermaid Lane, Philadelphia, Pennsylvania 19118. He received his AB in chemistry from the University of North Carolina at Chapel Hill in 1968 and his PhD in inorganic chemistry from Northwestern University in 1974. In 1978, Byler was awarded a National Research Council-National Academy of Sciences Postdoctoral Fellowship to work at USDA-ERRC and became a permanent member of the scientific staff the following year. Byler's current research includes spectroscopic studies of soil components as well as proteins. He is coauthor of more than 25 publications and is a member of Sigma Xi, the Royal Society of Chemistry, ACS, SAS, the Spectroscopy Society of Canada, and the Coblentz Society.*

*Janine N. Brouillette received her BS in chemistry from St. Joseph's University in 1984. While an undergraduate, she worked as a cooperative education student at USDA-ERRC. After graduating, she joined the scientific staff of the USDA-ERRC as a chemist in the plant and soil biophysics research group.*

*Heino Susi is a native of Estonia and completed his undergraduate chemistry education in 1950 at the Technische Hochschule of Aachen in the Federal Republic of Germany. After coming to the United States, he obtained his MS at the University of Oklahoma in 1952 and his PhD in physical chemistry at Purdue University in 1955. He spent one year at Rohm and Haas in Philadelphia before joining the scientific staff at USDA-ERRC, where he is now a senior scientist in the macromolecular and cell structure research group. Susi has published nearly 70 papers on spectroscopic research of a variety of biomolecules. He is a member of SAS, the Coblentz Society, and the Biophysical Society.*

## Protein structure by FTIR self-deconvolution

D. Michael Byler and Heino Susi

Eastern Regional Research Center, U. S. Department of Agriculture  
600 East Mermaid Lane, Philadelphia, PA 19118 USAAbstract

Fourier self-deconvolution was applied to the peptide-carbonyl stretching vibration (amide I mode) of more than 20 globular proteins in deuterium oxide solution. This band, which usually exhibits little discernible fine structure, was thereby resolved into three to nine components. The individual components were assigned to protein segments consisting of extended chains, helices, and various turns and bends. The areas of the components were evaluated by Gauss-Newton iterative curve fitting with the assumption of Gaussian band shapes. Quantitative estimations regarding secondary structure were made by calculating the sum of the areas of the components associated with a particular conformation as a fraction of the total amide I band area. The results for helix content and for extended chain content are in good agreement with literature values obtained from X-ray data.

Introduction

Fourier self-deconvolution has been previously applied to the FTIR spectra on numerous proteins both in the solid state and in deuterium oxide solution.<sup>1</sup> Of all the characteristic protein bands, usually labeled amide I to amide VII, the amide I band is the most useful for extracting information regarding the secondary structure (conformation).<sup>2</sup> Until recently, only qualitative information could be obtained<sup>2-6</sup>, although some quantitative studies have been carried out by curve fitting with the help of digital computers.<sup>7</sup> The present investigation was undertaken to determine if more detailed quantitative information can be obtained by combining Fourier self-deconvolution of the amide I band with an iterative curve fitting procedure.

Experimental

The proteins were dissolved in D<sub>2</sub>O with 0.01 N NaCl at concentrations of about 5% w/v and equilibrated until total exchange was achieved, usually about 24 hrs. The completeness of exchange was judged by the disappearance of the amide II band around 1550 cm<sup>-1</sup>. The use of D<sub>2</sub>O is necessary because the strong HOH bending mode at ca. 1640 cm<sup>-1</sup> overlaps with the amide I mode (1620-1700 cm<sup>-1</sup>) and makes sufficiently accurate solvent subtraction impossible<sup>1</sup>. FTIR spectra from 1600-1700 cm<sup>-1</sup> were measured with a Nicolet 7199 instrument. All spectra were obtained from 4000 co-added interferograms. Nominal instrument resolution was 2 cm<sup>-1</sup>; the cell pathlength was 0.075 mm. Deconvolution was carried out with a program supplied by Nicolet which is based on the algorithm of Kauppinen et al.<sup>8</sup> The parameters VFO and VF1 (which correspond to 2 X sigma and K in the original work<sup>8</sup>) were set at 13 cm<sup>-1</sup> and 2.4 for all samples. We emphasize that overdeconvolution must be carefully avoided if quantitative information is to be obtained from deconvolved spectra. Band centers were affirmed by second derivative spectroscopy.<sup>8,9</sup> Curve fitting was accomplished with the program ABCUS written in this Laboratory by W. C. Damert. Gaussian bandshapes were assumed for the deconvolved components. There is, at present, no theoretical justification for this approximation, but it does lead to quite satisfactory results. Iterative curve fitting did not always lead to a unique solution, as judged by a minimum RMS deviation between the calculated and observed data.

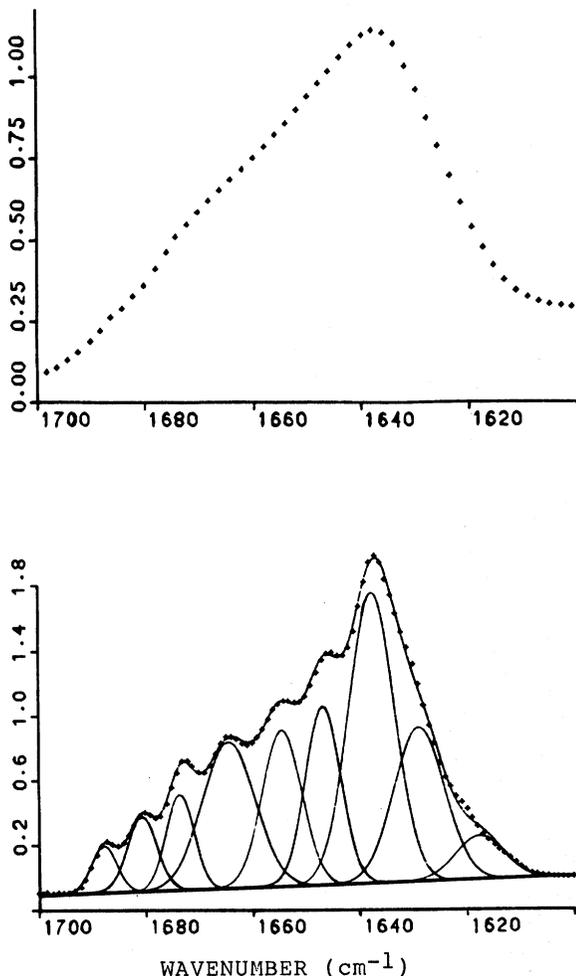


Figure 1.  $\alpha$ -chymotrypsin: original (top) and deconvolved (bottom) spectra.

"False solutions" can, however, usually be easily detected and rejected, because they involve unreasonably large values for one or more bandwidths.

### Results and discussion

Figure 1 shows the original spectrum, the deconvolved spectrum, and the fitted components for the amide I band of the protein  $\alpha$ -chymotrypsin. The RMS error is less than 0.5% of the maximum absorbance value. Table 1 gives the assignment of the observed components and the areas as fractions of the total amide I area. The assignments are based on previous work in this Laboratory<sup>1,2,5,7</sup> and on normal coordinate calculations by Krimm and coworkers.<sup>10,11</sup> Table 2 gives the assignment of amide I components based on the study of 24 globular proteins in our Laboratory. Table 3 gives the percentage of helix content and extended chain content for 5 proteins as calculated from data analogous to the ones given in Table 1. It is not possible to give similar comparisons for bends and turns, because these are not defined in a uniform manner by various authors. The work of Levitt and Greer<sup>12</sup> was selected for for X-ray data because, to the best of our knowledge, it represents one of the most consistent interpretations of crystallographic data in terms of conformation and covers a large number of proteins.

Table 1. Deconvolved FTIR spectra of  $\alpha$ -chymotrypsin in the amide I region (1620-1700  $\text{cm}^{-1}$ )

Wavenumber	Assignment	%Total Area
1688	Turns	2.3
1680	Turns	5.7
1673	Ex. Chain*	5.4
1665	Turns	15.5
1655	Helix	11.7
1647	Unordered	15.3
1638	Ex. Chain	25.7
1630	Ex. Chain	18.4

\*Extended chain.

Table 2. General assignment of amide I components

Wavenumber	Assignment
1624 $\pm$ 4	Ex. Chain*
1632 $\pm$ 2	Ex. Chain
1638 $\pm$ 2	Ex. Chain
1644 $\pm$ 3	Unordered
1654 $\pm$ 3	Helix
1663 $\pm$ 4	Turns
1675 $\pm$ 4	Ex. Chain
1683 $\pm$ 2	Turns
1688 $\pm$ 2	Turns
1694 $\pm$ 2	Turns

\*Extended chain.

Table 3. Protein secondary structure by FTIR and X-ray

Protein	% helix		% extended chain	
	FTIR	X-ray <sup>12</sup>	FTIR	X-ray <sup>12</sup>
Carboxypeptidase	40	39	33	30
$\alpha$ -chymotrypsin	12	10	50	49
Concanavalin A	4	2	60	60
Lysozyme	41	45	21	19
Papain	27	29	32	29

As seen in Table 3, the agreement between our FTIR values and values derived from X-ray crystallography is quite good. We are conscious of the fact that in Table 3 solution values are compared with solid state values. This appears unavoidable because no comprehensive set of secondary structure data for proteins in aqueous solution appears to exist. Nevertheless, we do need a set of some kind to establish the validity of the FTIR approach. It can not be emphasized too strongly that values for "% helix" or "% extended structure" are always somewhat subjective and depend on the manner in which they have been defined by the various investigators. With this in mind, the agreement between the presented X-ray values and our FTIR values must be regarded as quite good. The five proteins in Table 3 were selected out of a set of 24 so as to cover a reasonably broad range of both helix content and extended chain content. The presented accuracy and agreement with X-ray data is typical for the entire set of proteins studied.

Acknowledgement: We thank Janine Brouillette and Howard Fertman for obtaining the spectra.

### References

1. Yang, W.J., Griffith, P.R., Byler, D.M. & Susi, H., Appl. Spectrosc. 39, 282-287. 1985.
2. Susi, H., Meth. Enzymol. 26, 455-472. 1972.
3. Miyazawa, T., J. Chem. Phys. 32, 1647-1652. 1960.
4. Krimm, S., J. Mol. Biol. 4, 528-540. 1962.
5. Timasheff, S.N., Susi, H. & Stevens, L., J. Biol. Chem. 242, 5467-5473. 1967.
6. Elliott, A., & Ambrose, E.J., Nature 165, 921-922. 1950.
7. Rüegg, M., Metzger, V. & Susi, H., Biopolymers 14, 1465-1471. 1975.
8. Kauppinen, J.K., Moffatt, D.J. Mantsch, H.H. & Cameron, D.G. Appl. Spectrosc. 35, 271-276. 1981; Anal. Chem. 53, 1454-1457. 1981; Appl. Optics 20, 1866-1879. 1981.
9. Susi, H., & Byler, D.M., Biochem. Biophys. Res. Comm. 115, 391-197. 1983.
10. Krimm, S. & Bandekar, J. Biopolymers 19, 1-29. 1980.
11. Dwivedi, A.M. and Krimm, S., J. Phys. Chem. 88, 620-627. 1984.
12. Levitt, M., & Greer, J., J. Mol. Biol. 114, 181-239. 1977.