

MACROMOLECULAR CONFORMATION: SPECTROSCOPY

FOURIER TRANSFORM INFRARED SPECTROSCOPY

and other methods have confirmed the presence of 3Fe centers in *Azotobacter vinelandii* ferredoxin I,⁶³ *E. coli* nitrate reductase,⁶⁴ succinate dehydrogenase,⁶⁵ and *E. coli* fumarate reductase.⁶⁶ The identification, with the aid of low temperature MCD, that the latter two enzymes have 3 different FeS clusters, one each of 2Fe, 3Fe, and 4Fe, provides the first rational picture of the clusters in these enzymes clarifying much of past data.

⁶³ T. V. Morgan, P. J. Stephens, F. Devlin, B. K. Burgess, and C. D. Stout, *FEBS Lett.* **183**, 206 (1985).

⁶⁴ M. K. Johnson, D. E. Bennett, J. E. Morningstar, M. W. Adams, and L. E. Mortenson, *J. Biol. Chem.* **269**, 5456 (1984).

⁶⁵ T. B. Singer and M. K. Johnson, *FEBS Lett.* **199**, 189 (1985).

⁶⁶ J. E. Morningstar, M. K. Johnson, G. Cecchini, B. A. Ackrell, and E. B. Kearney, *J. Biol. Chem.* **260**, 13631 (1985).

with practical applications, including seven published in this series.⁶⁻¹⁰ The latest one, in 1982, concerns measurement of peptide hydrogen exchange in rhodopsin.

Polypeptides and proteins exhibit a total of nine characteristic absorption bands in the infrared region. These are usually termed the amide A, B, and amide I-VII bands.^{2,58} The amide I ($\sim 1630-1690 \text{ cm}^{-1}$) band has been found to be the most useful for protein structure studies by infrared spectroscopy.^{3-6,10-11c} For deuterated proteins the designations amide I', II', etc., are employed.⁸

For proteins—as distinct from many synthetic polypeptides—each characteristic absorption band is generally a composite, consisting of overlapping components representing α -helical segments, β -sheet sections, turns, and unordered regions.^{12a-15} These subbands usually cannot be resolved by conventional spectroscopic techniques because their inherent widths are greater than the instrumental resolution. Infrared spectroscopy, until recently,^{14a-s} has therefore been essentially a qualitative

⁶ D. L. Wood, this series, Vol. 4 [3], p. 104.

⁷ W. P. Jencks, this series, Vol. 6 [125], p. 914.

⁸ H. Susi, this series, Vol. 26 [117], p. 381.

⁹ H. Susi, this series, Vol. 26 [22], p. 455.

¹⁰ S. N. Timashoff, H. Susi, and J. A. Rupley, this series, Vol. 27 [23], p. 548.

¹¹ D. F. H. Wallach and A. R. Oseroff, this series, Vol. 33 [22], p. 247.

¹² H. B. Osborne and E. Nabedryk-Viala, this series, Vol. 38 [81], p. 676.

¹³ N. Miwa, *J. Anibior.* **35**, 1553 (1982).

¹⁴ R. J. Jakobsen, L. L. Brown, T. B. Hutson, D. J. Fink, and A. Veis, *Science* **220**, 1288 (1983).

¹⁵ R. Mendelsohn, G. Anderle, M. Jaworsky, H. H. Mansisch, and R. A. Dluhy, *Biochim. Biophys. Acta* **775**, 215 (1984).

¹⁶ R. M. Gendreau, *Proc. SPIE Int. Soc. Opt. Eng.* **553**, 4 (1985).

¹⁷ K. B. Smith, C. A. Penkowski, and R. J. Jakobsen, *Proc. SPIE Int. Soc. Opt. Eng.* **553**, 178 (1985).

¹⁸ V. E. Kotliansky, M. A. Glukhova, M. V. Bejanian, V. N. Smirnov, V. V. Filimonov, O. M. Zalic, and S. Yu. Venyaminov, *Eur. J. Biochem.* **119**, 619 (1981).

¹⁹ S. Yu. Venyaminov, M. L. Metciss, M. A. Chernousov, and V. E. Kotliansky, *Eur. J. Biochem.* **135**, 485 (1983).

²⁰ H. Susi and D. M. Byler, *Biochem. Biophys. Res. Commun.* **115**, 391 (1983).
²¹ J. M. Purcell and H. Susi, *J. Biomed. Biophys. Methods* **9**, 193 (1984).

²² W. J. Yang, P. R. Griffiths, D. M. Byler, and H. Susi, *Appl. Spectrosc.* **39**, 382 (1985).

²³ H. Susi, D. M. Byler, and J. M. Purcell, *J. Biomed. Biophys. Methods* **11**, 235 (1985).

²⁴ D. M. Byler and H. Susi, *Proc. SPIE Int. Soc. Opt. Eng.* **553**, 289 (1985).
²⁵ D. M. Byler and H. Susi, *Biopolymers* **25**, 469 (1986).

²⁶ D. M. Byler, J. N. Broutelle, and H. Susi, *Spectroscopy* **1**, 29 (1986).

²⁷ S. Krimm and Y. Abe, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2788 (1972).
²⁸ S. Krimm and J. Bandekar, *Biopolymers* **19**, 1 (1980).
²⁹ Yu. N. Chirgadze, O. V. Fedorov, and N. P. Trushina, *Biopolymers* **14**, 679 (1975).

[113] Resolution-Enhanced Fourier Transform Infrared Spectroscopy of Enzymes

By HEINO SUSI and D. MICHAEL BYLER

General Background

Infrared spectroscopy constitutes one of the oldest methods for studying the secondary structure of polypeptides and proteins. As early as 1950, before any detailed X-ray results were available, let alone circular dichroism or optical rotatory dispersion data, Elliott and Ambrose showed that the "amide I" band is observed around $1650-1660 \text{ cm}^{-1}$ for the α -helical conformation and around $1630-1640 \text{ cm}^{-1}$ for β -strands.¹ Since then basic theoretical work on the subject has been carried out primarily by Miyazawa and co-workers^{2,3} and by Krimm and his colleagues.^{4-8b} A number of reviews and summaries have been concerned

¹ A. Elliott and E. J. Ambrose, *Nature (London)* **426**, 921 (1990).

² T. Miyazawa, T. Shimanouchi, and S. Mizushima, *J. Chem. Phys.* **24**, 408 (1956).
³ T. Miyazawa, *J. Chem. Phys.* **32**, 1647 (1960).

⁴ S. Krimm, *J. Mol. Biol.* **4**, 528 (1962).

⁵ S. Krimm and Y. Abe, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2788 (1972).
⁶ S. Krimm and J. Bandekar, *Biopolymers* **19**, 1 (1980).
⁷ S. Krimm and O. V. Fedorov, and N. P. Trushina, *Biopolymers* **14**, 679 (1975).

tool for conformational studies of proteins, although computerized techniques did make some semiquantitative estimates possible.^{11,15,16}

The use of Fourier transform infrared spectroscopy (FTIR) has led to major improvements in this regard. In principle, FTIR provides several advantages over conventional dispersive techniques^{17a-c}: higher (1) resolution, (2) sensitivity, (3) signal-to-noise ratio (S/N), and (4) frequency expense of the other two. For protein structure studies, high sensitivity makes it possible to acquire usable infrared spectra of aqueous solutions;^{18a-c} such spectra are always notoriously difficult to obtain. The improved S/N ratio facilitates resolution enhancement of observed protein spectra through the application of (1) second derivative^{14a-h,21} and (2) Fourier self-deconvolution^{12a,14b-2} techniques. The latter is of particular importance because it offers the possibility of obtaining quantitative information on the conformation of proteins from infrared spectra.^{14c-e} Figure 1 shows (1) the original, (2) the deconvolved, and (3) the second derivative spectra of bovine α -chymotrypsin in D₂O solution from 1250 to 1800 cm^{-1} .^{14e} The strong spectroscopically unresolvable amide I' band, centering around 1640 cm^{-1} in the original spectrum, is resolved into seven components by both deconvolution and second derivative techniques. This newly discovered fine structure reflects different conformational entities. The sections which follow describe the basic principles, techniques, and applications of these methods in more detail.

Basic Theory

Fourier transform infrared spectrometers are fundamentally different in construction and operation than conventional dispersive instruments.

- ¹⁶ M. Rüegge, V. Metzger, and H. Susi, *Biopolymers*, **14**, 1465 (1975).
- ^{17a} P. R. Griffiths, in "Analytical Applications of FT-IR to Molecular and Biological Systems" (J. R. Durig, ed.), p. 11. Reidel, Dordrecht, 1980.
- ^{17b} P. R. Griffiths, *Science*, **222**, 297 (1983).
- ¹⁸ J. E. Bertie, in "Vibrational Spectra and Structure" (J. R. Durig, ed.), Vol. 14, p. 221.
- ¹⁹ J. L. Koenig, *Adv. Polym. Sci.*, **54**, 87 (1984).
- ²⁰ L. D'Esposito and J. L. Koenig, in "FTIR Spectroscopy" (J. F. Ferraro and L. J. Basile, ed.), Vol. 1, p. 61. Academic Press, New York, 1978.
- ²¹ J. L. Koenig and B. L. Tabb, in "Analytical Applications of FT-IR to Molecular and Biological Systems" (J. R. Durig, ed.), p. 241. Reidel, Dordrecht, 1980.
- ²² R. M. Gendreau, S. Winters, R. I. Leininger, D. Fink, C. R. Hassler, and R. J. Jakobsen, *Appl. Spectrosc.*, **35**, 353 (1981).
- ²³ S. Winters, R. R. Gendreau, R. I. Leininger, and R. J. Jakobsen, *Appl. Spectrosc.*, **36**, 404 (1982).
- ²⁴ M. Thétrein, M. Lafleur, and M. Pérotet, *Proc. SPIE Int. Soc. Opt. Eng.*, **553**, 173 (1985).
- ²⁵ D. C. Lee, D. A. Elliott, S. A. Baldwin, and D. Chapman, *Biochem. Soc. Trans.*, **13**, 684 (1985).

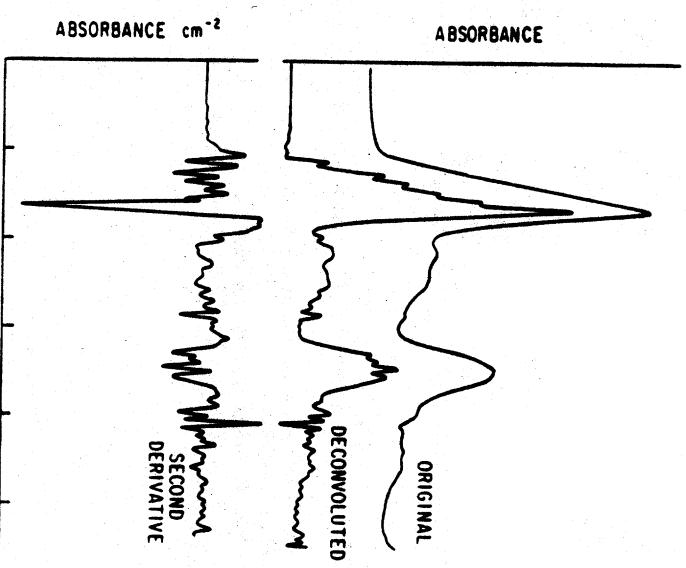


FIG. 1. FTIR spectrum of α -chymotrypsin in D₂O solution. Pathlength, 0.075 mm. Components used for deconvolved spectrum: $\sigma = 6.5 \text{ cm}^{-1}$, $K = 2.4$. Second derivative of original spectrum, obtained by Eq. (1) (Byler *et al.*^{14e}).

ments.^{17a-19} In the latter, a grating or prism disperses a collimated beam of infrared light onto a slit which effectively blocks all but a narrow range of frequencies from reaching the detector. By continuously changing the angle of the grating with respect to the incident light beam, a complete spectrum can be scanned, one spectral resolution element at a time. The FTIR instrument, by contrast, is nondispersive and makes use of an interferometer to encode data from the whole spectral range simultaneously. In general, the interferometer is some variation of the original design by Michelson.²² (A few commercial instruments now employ a refractively scanned interferometer,¹⁷ which differs markedly from the Michelson interferometer.)

Figure 2 depicts the principal features of a typical Michelson interferometer. Set at right angles to one another is a pair of plane mirrors, F and

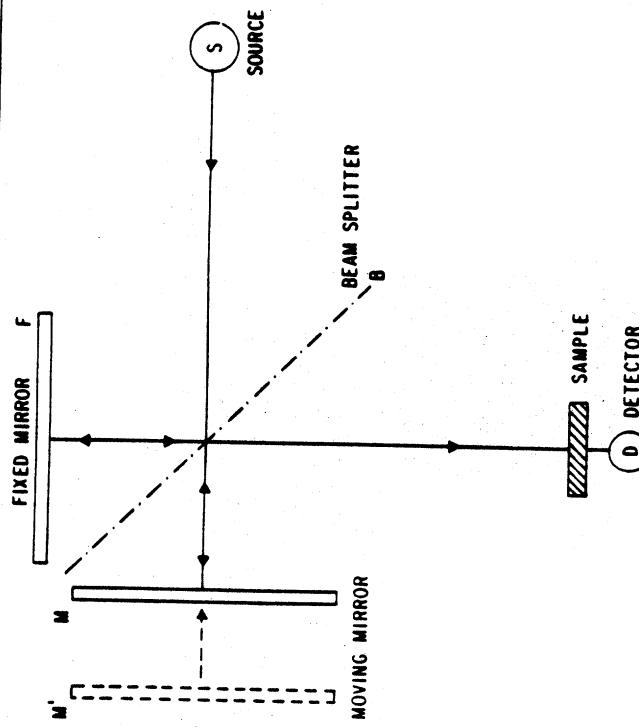


FIG. 2. Schematic representation of a Michelson interferometer.

M' , mounted between is a semireflecting film, or beamsplitter B, with its plane at a 45° angle to the mirror faces. As a collimated beam of light from the source S impinges on the beamsplitter B, half is reflected to mirror F and half is transmitted to mirror M. After reflection at the mirrors, the two beams reconverge at the beamsplitter B, where again each is 50% reflected and 50% transmitted. For simplicity, first consider a monochromatic light source of wavelength λ . Twice the difference in the distances from the beamsplitter B to each of the two mirrors is designated as x , the optical retardation or optical path difference. If x is zero or an integral multiple of λ , the two beams will recombine at B in-phase. Due to constructive interference, the signal at the detector will be of maximum intensity. On the other hand, if x has any other value, the two beams of light will be partially out-of-phase, resulting in destructive interference and decreased detector signal. When $x = (n + 1/2)\lambda$ ($n = 0, 1, 2, \dots$), the signal is zero. If mirror F is fixed in position while mirror M moves at a constant velocity v through some distance r , the signal observed at the detector will be a cosine wave whose frequency is represented by $f = v/\lambda$. (Here λ is the wavenumber frequency of the incident radiation.)

The amplitude or intensity of this signal as a function of x is $I(x)$ and is called an interferogram. $I(x)$ is proportional to $\cos(2\pi kx)$.

In the case of polychromatic light, all frequencies will be in-phase only at zero path difference ($x = 0$). At all other values of x , varying degrees of destructive interference will occur. Now the interferogram resulting from one sweep of the moving mirror M will be proportional to the sum of the cosine waves, $\sum_i A_i \cos(2\pi k_i x)$. A_i is the maximum amplitude of the cosine for each incident frequency, k_i . (For a frequency continuum, i becomes infinite and the sum is replaced by an integral.) This interferogram $I(x)$ now has maximum amplitude at $x = 0$. If certain frequencies of the incident radiation are absorbed by a sample, the interferogram changes because the amplitudes of the cosine waves of the absorbed frequencies decrease. Even when such changes are readily apparent upon visual inspection of the interferogram they are difficult to interpret. The interferometric data from the "time domain" are therefore Fourier transformed into the "frequency domain" to give an uncorrected, "single-beam" absorption spectrum of the sample, $E(k)$. Dividing $E(k)$ by the spectrum of the incident beam with no sample present, $E(k_0)$, one obtains the spectrum of the sample in percentage transmittance. This is commonly called a "ratioed" spectrum.

FTIR spectrometry offers several theoretical advantages over dispersive infrared measurements¹⁷⁻¹⁹:

1. Because the throughput of the incident light is not slit limited, FTIR spectrometers are inherently more sensitive than dispersive instruments (*Jacquinot's advantage*). For a given source more energy will reach the detector, resulting in a greater signal-to-noise (S/N) ratio. (This advantage is somewhat offset because dispersive instruments can use slow-response thermocouple detectors which have higher detectivities than the most sensitive, low-temperature FTIR detectors for the mid-infrared region.)
2. FTIR spectrometers simultaneously encode all spectral frequencies to give a complete spectrum in a matter of seconds (*Fellgett's or multiplex advantage*). Because the S/N ratio is proportional to the square root of the number of scans, signal averaging the data from a large number of scans can significantly increase the S/N ratio within a relatively short time. In addition, rapidly occurring changes may be monitored rather easily.
3. Modern, high-speed FTIR spectrometers contain a laser reference interferometer to facilitate digitization of data and for frequency calibration. Because the laser frequency is known to at least seven significant figures, its interferogram frequency is used to calibrate the frequencies of the digitized infrared data points to better than 0.01 cm^{-1} (*Comme's advantage*).

large). The peak positions of observed infrared bands of condensed-phase samples generally cannot be measured with such accuracy, particularly if the bands overlap or are broad. Use of Fourier self-deconvolution and of second derivatives, however, can increase substantially the accuracy of the measured peak positions.

4. Finally, because such good S/N ratios and high wavenumber precision are possible, FTIR spectra can be manipulated easily and efficiently by a computer. This facilitates such mathematical data treatments as precise measurement of peak maxima,²¹ interactive spectral subtraction,²² calculation of second derivatives,^{14a-h,21} Fourier self-deconvolution,^{12a,14b-g} and curve fitting.^{14d-g}

One additional distinction between FTIR and dispersive spectroscopy is worth noting. For the former, resolution remains constant across the spectral range, but the S/N ratio does not. In dispersive spectrometry, just the reverse is generally true: constant S/N can be achieved, but only at the expense of varying resolution.

Several recent reviews give more detailed treatment of FTIR theory and instrumentation, as well as additional background references.^{17a-19}

Second Derivative Spectra

Because FTIR spectra are digitally encoded with one data point n every $\Delta k/(2^m) = \Delta W$ frequency units, where Δk is the nominal instrumental resolution selected (in cm^{-1}) and m is the number of times the interferogram is zero-filled prior to Fourier transformation, the second derivative of the spectrum may be calculated by a modification of a straightforward analytical method described by Butler and Hopkins.^{24a} In particular, at data point n the value of the second derivative A''_n in absorbance units/wavenumber² is

$$A''_n = (A_{n+1} - 2A_n + A_{n-1})/(\Delta W)^2 \quad (1)$$

where A_n is the absorbance at data point n of the original spectrum. (Note that this function gives the second derivative spectrum without any frequency measurement; we occasionally employ for this purpose a seven- or a nine-point Savitsky-Golay function.^{24b}

The intrinsic shape of a single infrared absorption line of an isolated

molecule may be approximated by a Lorentzian function.^{25,26} The molecule may be approximated by a Lorentzian function.^{25,26} The frequency of the original band center, k_0 . The half width of the second derivative, σ'' , is related to the half width of the original line^{26a-27} by Eq. (2b) into Eq. (1) gives the second derivative of this function:^{25a-26b}

$$A'' = \frac{-2B(1 - 3Bk'^2)}{\sigma\pi(1 + BK'^2)} \quad (2b)$$

where A is the absorbance, 2σ is the width at half height, $k' (= k - k_0)$ is the frequency referred to the band center at k_0 , and $B = 1/\sigma^2$. Substituting Eq. (2b) into Eq. (1) gives the second derivative of this function:^{25a-26b}

$$A'' = \frac{-2B(1 - 3Bk'^2)}{\sigma\pi(1 + BK'^2)} \quad (3)$$

and the peak intensity of the second derivative, A''_0 , to that of the original intensity A_0 by

$$A''_0 = -2A_0/\sigma^2 \quad (5)$$

Eq. (5) is derived by dividing Eq. (3) by Eq. (2) and noting that $k' = 0$ at the band center. The peak height of the second derivative is thus proportional to the original peak height and inversely proportional to the square of the original half width. Thus weak, but sharp lines, such as arise from water vapor, noise, or interference fringes are greatly accentuated relative to the much broader lines of the condensed phase sample. For real spectra with overlapping bands which deviate from Lorentzian shape, the relationships are more complex. Nonetheless, the above formulas do provide a reasonable approximation for interpreting second derivative spectra.

The original method of Butler and Hopkins for analytically determining second derivative spectra^{21,25a-26b} was to take the first derivative of the

^{24a} W. F. Maddams and W. L. Mead, *Spectrochim. Acta* **38A**, 437 (1982).

^{24b} W. F. Maddams and M. J. Southon, *Spectrochim. Acta* **38A**, 459 (1982).

²⁵ J. K. Kauppinen, D. J. Moffatt, H. H. Mansch, and D. G. Cameron, *Appl. Spectrosc.* **35**, 271 (1981).

²⁶ J. K. Kauppinen, D. J. Moffatt, H. H. Mansch, and D. G. Cameron, *Anal. Chem.* **53**, 145 (1981).

²⁷ J. K. Kauppinen, *Spectrom. Techn.* **3**, 199 (1983).

²⁸ H. H. Mansch, H. L. Casal, and R. N. Jones, *Adv. Spectrosc.* **13**, in press (1986).

- ²³ D. G. Cameron, J. K. Kauppinen, D. J. Moffatt, and H. H. Mansch, *Appl. Spectrosc.* **36**, 25 (1982).
- ²⁴ W. L. Butler and D. W. Hopkins, *Photochim. Photobiol.* **12**, 439, 452 (1970).
- ²⁵ A. Savitsky and M. J. E. Golay, *Anal. Chem.* **36**, 1627 (1964).

original spectrum twice, using the function

$$A_n' = [A_{n+1} - A_{n-1}] / \Delta W \quad (6)$$

This method, however, results in a second derivative function which differs somewhat from Eq. (1):

$$A_n'' = (A_{n+2} - 2A_n + A_{n-2}) / (\Delta W)^2 \quad (7)$$

In this case, the second derivative spectrum is smoothed compared to that calculated by Eq. (1). Smoothing occurs because now the calculation of A_n'' spans five data points including A_n , whereas in Eq. (1) only three data points are involved.

Savitsky and Golay have described another approach which combines smoothing and derivatization.^{24b} Here the spectral curve is fitted to a polynomial by a least-squares calculation. With the polynomial coefficients in hand, derivatization is straightforward. One additional method for obtaining derivative spectra involves the use of Fourier transformation,^{24a-27,28a,b} again with some simultaneous smoothing.

Fourier Self-Deconvolution

The Fourier transformation of a Lorentzian band, $E(k)$ (Fig. 3A), of half-width at half-height, σ (cm^{-1}), results in a time-domain function, or interferogram, $I(x)$, which is an exponentially damped cosine wave (Fig. 3A).^{24a-27} The envelope of this function is a decay curve whose exponent is directly proportional to σ . If $I(x)$ is multiplied by the product of an apodization function $D(x)$ and an exponentially increasing weighting function, the exponent of the decay curve and thus the rate of decay of $I(x)$ will be decreased. Therefore, when the reverse Fourier transform is calculated, the width of the resulting spectral band, $E_1(k)$, will be less than that of the original undeconvolved band, $E(k)$ (Figs. 3B-D).

If the positive exponent of the weighting function has the same absolute value as the negative exponent in the decay function, the resulting interferogram $I_1(x)$ will be a cosine wave truncated at $x = L$. The value of L for $I_1(x)$ is determined by the point L where the apodization function $D(x)$ goes to zero. Instrument resolution, Δk , at which the original spectrum was measured places an approximate upper limit on L of $1/\Delta k$.^{27,28a,b} The reverse Fourier transform of $I_1(x)$ gives a sinc function ($\text{sinc } x = [\sin x]/x$) (Fig. 3B). The characteristic side-lobes found on either side of the central maximum of this function clearly show that resolution enhancement of bands beyond $L = 1/\Delta k$ is not possible, even for perfectly

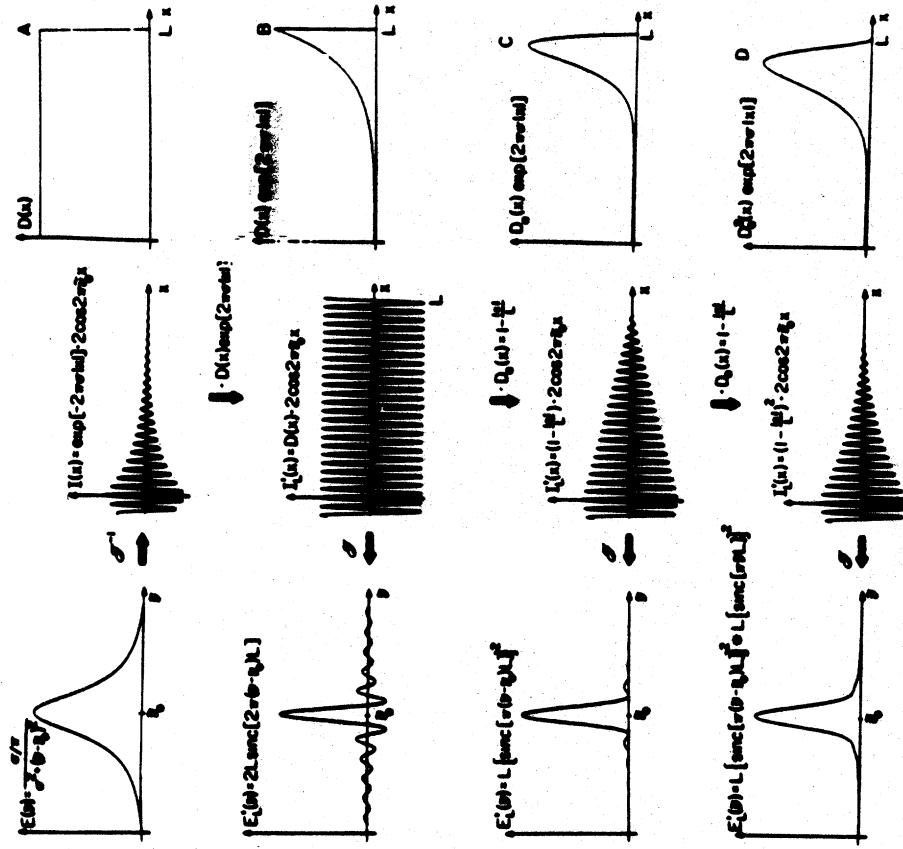


FIG. 3. Illustration of the various steps of the Fourier self-deconvolution procedure starting with a Lorentzian line at p_0 and using different apodization functions, $D_1(x)$; in (B), $D_1(x)$ equals $D_2(x)$, the "boxcar function" (see Eq. (6) of text); in (C), $D_2(x)$ equals $D_3(x)$, the "triangular function"; in (D), $D_3(x) = D_2(x)$. Except for (A), the middle column shows the interferograms, $I_1(x)$, resulting from the application of the functions, $D_1(x)$ $\exp(2\pi k x/L)$ (right-hand column), to the interferogram $I(x)$. These functions are all scaled to the same height. In the left-hand column, except for (A), the line shapes $E_1(k)$, resulting from the self-deconvolution are shown. Note: throughout the text, k is used for wavenumber instead of ν (Kuuppinen *et al.*^{28a}).

^{24a} W.-J. Yang and P. R. Griffiths, *Proc. SPIE Int. Soc. Opt. Eng.* **209**, 263 (1981).

^{24b} W.-J. Yang and P. R. Griffiths, *Comput. Enhanced Spectra*, **1**, 157 (1983).

noise-free data. Indeed, to avoid the appearance of side-lobes in the deconvolved spectrum, alternative apodization functions to the frequently used boxcar function,

$$D(x) = \begin{cases} 1, & \text{for } x \leq L \\ 0, & \text{for } x > L \end{cases} \quad (8)$$

are commonly used. These functions reduce the size of the side lobes, but at the cost of decreased resolution enhancement. Yang and Griffiths suggest as a rule-of-thumb that, at best, the full width at half-height of bands after deconvolution cannot be reduced to less than about $1.5(\Delta k)^{-1}$.^{29a,b}

The necessity of truncating the interferogram after Fourier transformation at $x = L$ tends to introduce new, nonrandom noise with a periodicity of about $1/L$.^{26a,29b} This periodic noise will appear across the whole spectrum, in contrast to the side-lobes mentioned above which are observed only along either side of an over-deconvolved band. With care in the choice of deconvolution parameters, such noise will have minimal intensity. Nonetheless, because it is generally present to some degree in all deconvolved spectra, one must firmly resist the temptation to attempt to obtain increased resolution enhancement of deconvolved data by attempting the second derivative of the deconvolved spectrum. Such a procedure will intensify the otherwise only faintly perceptible periodic noise with a concomitant loss in S/N. This, in turn, will make objective discrimination difficult, if not impossible.

FTIR Spectroscopy of Aqueous Solutions

Strong absorption by water throughout much of the mid-infrared spectral region has always made the acquisition of usable infrared data from aqueous solutions difficult. Studies of aqueous protein solutions are further complicated because the bending vibration of water absorbs strongly near 1640 cm^{-1} , right in the midst of the region ($1630\text{--}1690\text{ cm}^{-1}$) where the conformation-sensitive amide I protein vibrations occur. With noncomputerized dispersive instruments, extremely careful differential work was necessary to obtain any useful results.³⁰ The high frequency accuracy and reproducibility, and high S/N ratios, obtainable with FTIR instruments now make such measurements somewhat less difficult experimentally.^{12a,20a-c,21} Instead of using differential techniques with very carefully matched cells, digitized FTIR spectra of the solution and of the solvent are obtained separately, and the latter subtracted from the former. Neither subtraction nor differential procedures are as straightforward as they first appear. Whenever a solute is present, changes occur in the frequency, width, and height of the solvent band, even for weakly interact-

FOURIER TRANSFORM INFRARED SPECTROSCOPY
CHARACTERISTIC AMIDE I AND AMIDE II FREQUENCIES
(cm^{-1}) FOR PROTEINS IN H_2O SOLUTION AS MEASURED BY
FTIR SPECTROSCOPY

Protein	Predominant conformation	Amide I	Amide II
Hemoglobin	α -Helix ^a	16.66	1547
Ribonuclease A	β -Strands ^b	16.56	1548
β -Lactoglobulin A	β -Strands ^c	1632	1551
α -Casein	Unordered ^d	16.55	1551
Koehig and Tabb, ^{30b} Levitt and Greer, ³¹ Sawyer <i>et al.</i> ¹¹			
H. Susi, S. N. Timashoff, and L. Stevens, <i>J. Biol. Chem.</i> 242, 5460 (1967) and S. N. Timashoff, H. Susi, and L. Stevens, <i>J. Biol. Chem.</i> 242, 5467 (1967).			
Frequencies too high for a typical β -protein.			

ing, low-polarity solvent systems. For water and other polar, hydrogen-bonding solvents, such effects become much more pronounced. When

the protein amide I band and the bending mode of water, the residual features of the solvent spectrum which remain after subtraction often distort the solute bands so that accurate measurement of their frequency and intensity is no longer possible. Because absorption by water is significantly lower in the amide II region ($1530\text{--}1560\text{ cm}^{-1}$), only minimal difficulty is encountered with solvent subtraction for these protein bands.

Side-chain bands, associated with CH_2 , CH_3 , COO^- , NH_2 , and SH groups can also be easily observed in H_2O solution; the latter,³⁰ though very weak, falls within the relatively clear spectral region around 2560 cm^{-1} .

Koenig and Tabb^{20b} have tabulated a number of observed infrared frequencies for selected proteins in H_2O solution. Table I lists some of these values. Although the amide II frequencies can be observed with considerable accuracy (because of minimal interference by water), all appear within 4 cm^{-1} of one another and thus provide no reliable conformational correlations. Frequency shifts of the amide II band accompanying dissolution, as reported by these authors,^{30b} are such that these bands become much less useful as a conformational indicator in H_2O solution than they are in the solid state.¹⁴ The amide I frequencies, on the other hand, are uncertain because of the difficulties inherent in compensating

for the strong water band about 1640 cm^{-1} . They agree well with earlier work in the case of β -lactoglobulin and α -casein,²⁵ but the hemoglobin frequency (1656 cm^{-1}) is 4 cm^{-1} higher than the old value for myoglobin,²⁶ although both proteins have an almost identical secondary structure for their helical segments.³¹ The frequencies given for ribonuclease seem altogether too high for a protein with a high β -structure content³¹ (see below).

Both the older differential technique and the computerized FTIR subtraction procedure are prone to serious errors for aqueous solutions in the region around 1640 cm^{-1} . Thus, we conclude that even with the improved accuracy and sensitivity of FTIR spectroscopy, conformational studies of proteins are best carried out in D_2O solution,¹⁴ where no strong solvent bands appear close to the amide I' frequency region.

Application of Second Derivative FTIR Spectra

The intrinsic shape of an ideal infrared absorption line is approximated by a Lorentzian function, as described in the theoretical section. In the second derivative spectrum of such a line the peak frequency is identical with the original peak frequency and the half width is reduced by a factor of 2.7 .^{26,27} The peak height is proportional to the original peak height (with opposite sign), and inversely proportional to the square of the original half-width [see Eq. (5)]. There are side-lobes on either side of the principal peak.^{25,26} The net result is a very much sharper line accompanied by side-lobes. For real spectra the relationships are more complex, but the above simplified relationships provide a good starting point for interpreting the second derivative spectra of proteins.^{14a,b,21} The main value of second derivative spectra lies in the ease with which the peak frequencies of unresolved components can be identified. For side-chain bands a considerable sharpening is also observed, which makes identification and assignments much easier. Quantitative information, however, is difficult to obtain from second derivative spectroscopy because of the complex patterns created by overlapping peaks and side-lobes.

The bottom curve in Fig. 1 depicts the second derivative FTIR spectrum of α -chymotrypsin.^{14a} The original unresolved strong amide I band around 1640 cm^{-1} now exhibits six sharp peaks. The increased resolution of the side-chain bands below 1600 cm^{-1} is also clearly evident. Figure 4 shows the second derivative spectra of two other proteins^{14a} which further illustrate the power of the method. The original FTIR spectra of these proteins are as nondescript as the top curve in Fig. 1.

In the second derivative spectrum of bovine hemoglobin (Fig. 4A)^{14a}

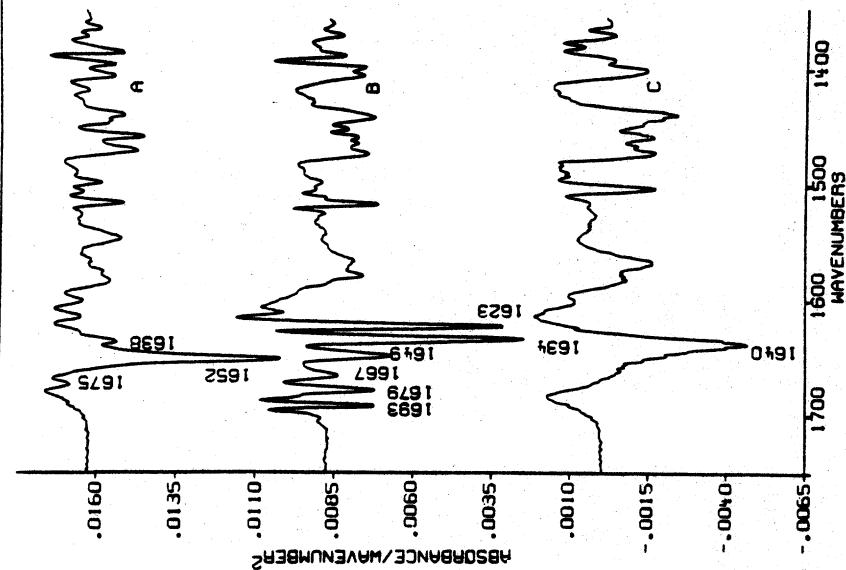


FIG. 4. Smoothed second derivative FTIR spectra in D_2O : (A) hemoglobin, pD 7; (B) native β -lactoglobulin A, pD 7; (C) denatured β -lactoglobulin A, pD 13; concentration 5% w/v (Susi and Byler¹⁴).

there is a single strong peak at 1652 cm^{-1} in the amide I region, obviously representing the α -helix (Table II). Hemoglobin is about 80% α -helical and contains no β -structure.³¹ The weak peaks at ~ 1638 and 1675 cm^{-1} can therefore be assigned to short extended chains connecting the helical cylinders.^{14f}

Figure 4B gives the second derivative spectrum of native bovine β -lactoglobulin A. Circular dichroism³² and infrared studies^{14f} suggest an α -helix content of 10–15% and a β structure content of about 50%. A recently published note on the crystal structure of this protein confirms

TABLE II

CHARACTERISTIC FREQUENCIES (cm⁻¹) AND
ASSIGNMENTS OF AMIDE I' COMPONENTS^a FOR 19
GLOBULAR PROTEINS

Mean frequency	Assignment
1624 ± 4 ^b	Extended chain
1631 ± 3	Extended chain
1637 ± 3	Extended chain
1645 ± 4	Unordered
1653 ± 4	Helix
1663 ± 4	Turns
1671 ± 3	Turns
1675 ± 5	Turns
1683 ± 2	Extended chain
1689 ± 2	Turns
1694 ± 2	Turns

^a Byler and Susi¹⁴; Byler *et al.*¹⁴.

^b Maximum range of the observed frequency for each amide I band component.

these results.³³ The strong 1634 cm⁻¹ peak is evidently associated with the second β structure band and with turns (Table II), as in bovine ribonuclease A.^{14a} It is interesting to note that the 1623 cm⁻¹ peak is absent in ribonuclease A but present in concanavalin A,^{14b} which also has a very high β content.³¹ Perhaps this peak represents a variation of the β structure present in Jack Bean concanavalin A and β -lactoglobulin A, but not in ribonuclease A. Further study is evidently required. Figure 4C gives the second derivative spectrum of denatured β -lactoglobulin A, which is assumed to be in a "random form."^{14a,b} All sharp peaks associated with helix, extended chain, and turns have disappeared.

Other spectral features are not as easily interpreted but are not related to the secondary structure. The weak peak around 1600 cm⁻¹ is too low for an amide I' component^{1,2}; it is probably caused by aromatic side chain groups.¹⁴ Other side chain groups of histidine, tryptophan, and phenylalanine, as well as the asymmetric stretching mode of side chain COO groups also absorb in the 1550 to 1610 cm⁻¹ region.¹⁴ The very stable, sharp 1515 cm⁻¹ band is associated with tyrosine residues.¹⁶ From ~1430 to 1480 cm⁻¹ we have overlapping bands caused by (1) CH₂ and CH₃,

bending modes of side chains, (2) the amide II' mode, essentially ND bending, and (3) bending modes of traces of HOD.^{14a,14} Between 1360–1380 cm⁻¹ the symmetric CH₃ bending vibrations of side chains are expected.³⁴ More detailed assignments must wait a thorough study of more proteins. It is evident, nevertheless, that second derivative spectra furnish new information about the side chains as well as the secondary structure of proteins.

Application of Fourier Self-Deconvolution and Curve Fitting

Figure 5 shows the self deconvolved FTIR spectra of ribonuclease A along with the original spectra. The deconvolved spectra are resolved into Gaussian components by means of a computer program which uses Gauss-Newton iteration. Deconvolution, as carried out by the described method,^{14,25a-28} requires two constants as computer input: the estimated half-width at half-height of the unresolved bands, σ , and the resolution bands, i.e., the "efficiency" factor, K , which reflects the narrowing of the unresolved constants is of utmost importance for obtaining meaningful half-width of the unresolved components as possible. When unresolved components with different half-widths must be deconvolved simultaneously, a compromise becomes necessary. This is frequently the case for protein spectra. A choice of $\sigma = 6.5$ has been found to be satisfactory,¹⁴ although lower values might be necessary for quantitative work. A maximum value for K can be estimated by the approximate relationship: $K_{\max} \approx \log(S/N)$. A K value of 2.4 would thus require an S/N ratio of at least 260. In practice, K values higher than 3 are rarely used.²⁹ It must be strongly emphasized that an improper selection of σ and/or K can lead to serious errors and artifacts. Deconvolution increases with increasing σ , but too large a value of σ leads to "overdeconvolution," resulting in side-lobes and distorted spectra. Too high a K value, in turn, leads to excessive nonrandom noise with a period of about $1/L$.^{26a,27} This, in turn, can distort the shape of weak bands and also introduce spurious peaks. Self-deconvolution, like second derivative spectroscopy, enhances noise, interference fringes, impurity bands, and water vapor bands, as well as the true bands of the sample. A high S/N ratio, high nominal instrument resolution, and well purged instruments are therefore essential for good results.

^a L. Sawyer, M. Z. Papiz, A. C. T. North, and E. E. Eliopoulos, *Biochem. Soc. Trans.*, 13,

265 (1985).

^b J. J. Bellamy, "The Infra-red Spectra of Complex Molecules," Vol. 1, 3rd Ed., pp. 6, 8, 21-27, 196-200, Chapman & Hall, London, 1975.

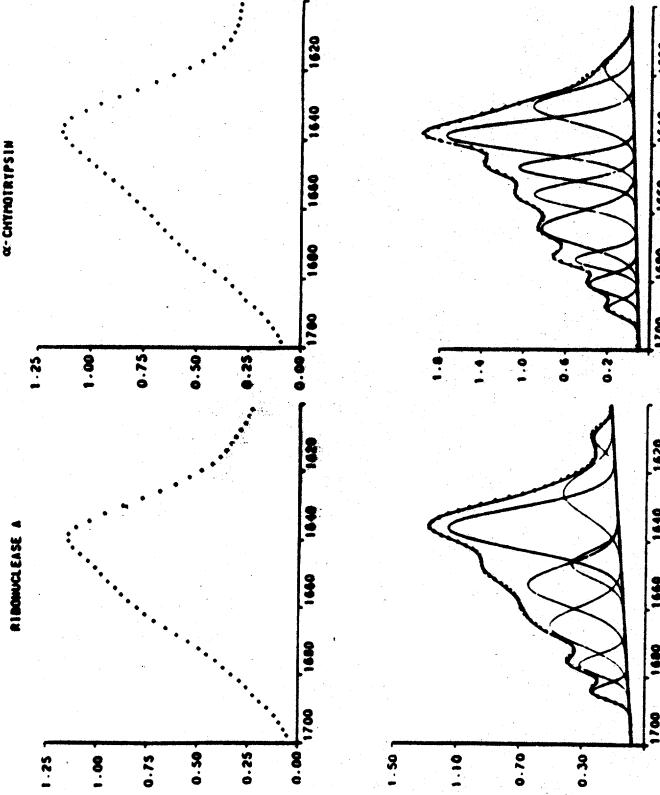


FIG. 5. Amide I' bands of ribonuclease A and α -chymotrypsin. Upper curve: digitized original FTIR spectrum (crosses), 5% w/v in D₂O, pathlength 0.075 mm. Lower curve: deconvolved spectrum (crosses) and individual Gaussian components (solid lines) (deconvolution constants: $\sigma = 6.5 \text{ cm}^{-1}$, $K = 2.4$). Component bands were calculated by least-squares fit using a Gauss-Newton iteration. Solid overall curve is obtained by summation of the components. Root mean square (RMS) error: 0.01 absorbance units (Byler *et al.*¹⁴).

The interpretation of deconvolved spectra¹⁴ (Fig. 5) is quite analogous to the interpretation of the second derivative spectra discussed in the previous section. Ribonuclease has an approximate α -helix content of 22% and a β -structure content of about 46%.³¹ The 1636 cm^{-1} band can be associated with β segments and the much weaker band near 1660 cm^{-1} with helical segments (Table II).^{30,14} The weak peaks between 1655 and 1700 cm^{-1} evidently correspond to the second β structure band and to turns. Comparison with the spectrum of bovine ribonuclease S,¹⁴ which has a very high β structure content and relatively few turns,³¹ suggests that the 1676 cm^{-1} band (Fig. 5) can be associated with the β structure while the remaining weak bands are due to turns.

In the spectrum of α -chymotrypsin the bands close to 1627, 1637, and 1674 cm^{-1} can be associated with β -strands, the band near 1653 cm^{-1} with

TABLE III
PERCENTAGE HELIX AND EXTENDED CHAIN BY FTIR AND X RAY FOR SIX TYPICAL PROTEINS

Protein	Helix (%)		Extended chain (%)	
	FTIR ^a	X ray ^b	FTIR ^a	X ray
Carboxypeptidase	40	39	33	30
α -Chymotrypsin	12	10	50	44
Concanavalin A	4	2	60	60
Lysozyme	41	45	21	19
Papain	27	29	32	29
Ribonuclease A	21	22	50	46

^a Byler and Susi.¹⁴

^b Levitt and Greer.³¹

the α -helix and the bands close to 1687, 1681, and 1665 cm^{-1} with turns (Table II).¹⁴⁻⁸

Figure 5 demonstrates that deconvolved amide I' spectra can be resolved into components with easily measurable areas. Herein lie the seeds for quantitative conformational analysis by FTIR spectroscopy.

Based on recent systematic studies of the frequency distribution of 138 observed amide I band components for 19 globular proteins with widely varying proportions of α -helix and β structures,¹⁴⁻⁸ a recurring pattern of bands at a limited number of distinct frequencies becomes apparent. As shown in Table II by the mean frequency value and narrow absolute range of each, these 11 characteristic amide I components show relatively little overlap with one another. The table also presents proposed assignments for each characteristic frequency to a particular type of secondary structure.¹⁴⁻⁸

Table III gives the percentage of helix content and extended chain content for six typical proteins as calculated from FTIR data.¹⁴⁻⁸ Corresponding values for the secondary structure of these proteins taken from Levitt and Greer's comprehensive examination of protein X-ray data³¹ are given for comparison. Their study represents one of the most thorough and consistent interpretations of protein crystallographic data in terms of conformation of which we are aware, and covers a total of 62 proteins. It is not possible to give similar comparisons for bends and turns, because these protein substructures do not have the same type of periodic, repeating structures found in helices and sheets. In addition, bends and turns usually consist of only a small number (four or less) of adjacent residues. For these reasons, whether a specific amino acid residue in a given pro-

tein belongs to a turn rather than the end of a helix or β -strand continues to elude precise definition.

As seen in Table III, the agreement between the FTIR results and values derived from X-ray crystallography is quite good. In Table III FTIR results from solution data¹⁴⁻¹⁸ are compared with values from solid state X-ray studies.³¹ At present, this is unavoidable because no uniform, comprehensive set of secondary structure data for proteins in aqueous solution has appeared in the literature.

One cannot emphasize too strongly that values for "percentage helix" or "percentage extended structure" are always somewhat subjective even when based on accurate bond lengths and angles. Ambiguity arises from the uncertainty in the choice of the exact point along the peptide chain where one segment of 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^{14-18,31} but also on just how regular any of these regions of protein substructure really are. With these uncertainties in mind, the agreement between the reported X-ray values³¹ and the FTIR values¹⁴⁻¹⁸ is particularly encouraging.

Despite such caveats and limitations, analysis of protein secondary structure in aqueous solution by judicious curve fitting of deconvolved FTIR spectra offers great promise for future studies of protein secondary structure including conformational changes which result when the biomolecule is subjected to changes in its environment.

Solvent Denaturation Studies

FTIR spectra, resolution enhanced by second derivative techniques and Fourier self-deconvolution, permit quite detailed conformational studies related to solvent denaturation.^{14,35} Figure 6 shows the deconvolved spectrum, the original FTIR spectrum, and the second derivative spectrum of native bovine chymotrypsinogen A in D₂O solution, and the corresponding spectra as obtained in 60% (v/v) O-deuterated methanol (MeOD) in D₂O. The spectrum of the native protein (Fig. 7A), which contains about 45% β structure and 11% α -helix,³² is somewhat similar to the ribonuclease spectrum shown in Fig. 5. β structure bands are observed at 1636 (strong) and 1674 (weak) cm^{-1} . The 1647 cm^{-1} band is probably associated with unordered sections and the remaining bands with turns (Table II).^{14,15,16} The denatured protein exhibits bands at 1617 cm^{-1} (strong) and 1686 cm^{-1} (weak). Quite similar results are obtained in isopropanol-*d* solution.^{14b} The bands at 1617 and 1686 cm^{-1} correspond to

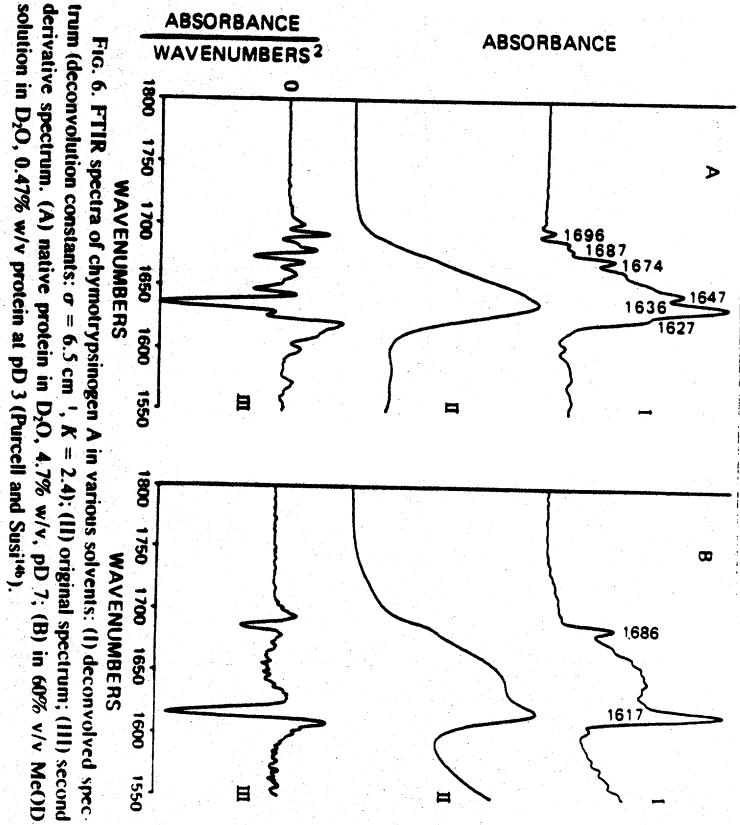


FIG. 6. FTIR spectra of chymotrypsinogen A in various solvents: (I) deconvoluted spectrum (deconvolution constants: $\sigma = 6.5 \text{ cm}^{-1}$, $K = 2.4$); (II) original spectrum; (III) second derivative spectrum. (A) native protein in D₂O, 4.7% w/v, pH 7; (B) in 60% v/v MeOD solution in D₂O, 0.47% w/v protein at pH 3 (Purcell and Susi^{14b}).

no conformation previously observed in the infrared spectra of globular proteins. They are, nevertheless, close to usually observed β structure bands and probably indicate the presence of extended strands of some kind. The ill-defined absorption between the two sharp bands could be associated with unordered hydrated segments and/or irregular helices.^{14b} The usefulness of FTIR spectroscopy as a tool for denaturation studies is further illustrated in Fig. 7, which shows the spectra of the native form and three different denatured forms of the protein β -lactoglobulin A. Figure 7A, the spectrum of the native material, is typical for a protein with a high β structure content; the assignment was discussed above in conjunction with the second derivative spectrum shown in Fig. 4B. Figure 7B shows the spectrum of alkaline-denatured β -lactoglobulin A at pH 13. A single strong, broad band is observed around 1640 cm^{-1} , suggesting hydrated unordered chains.^{14c} Figure 7C shows the spectrum obtained in acidic MeOD solution. Here one observes a strong band at 1647 cm^{-1} and two weaker ones at 1618 and 1687 cm^{-1} . The latter suggest an extended chain conformation similar to the one observed for chymotrypsinogen A

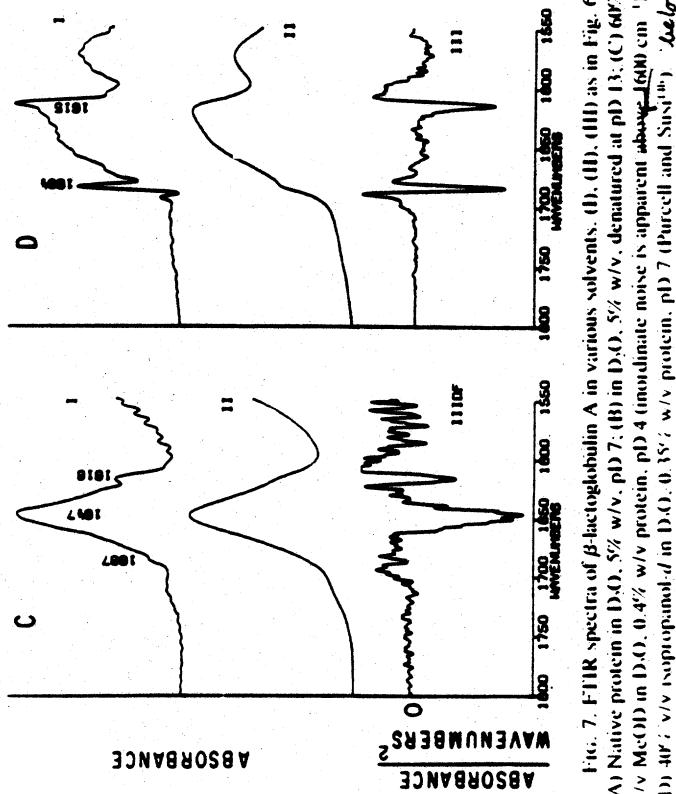
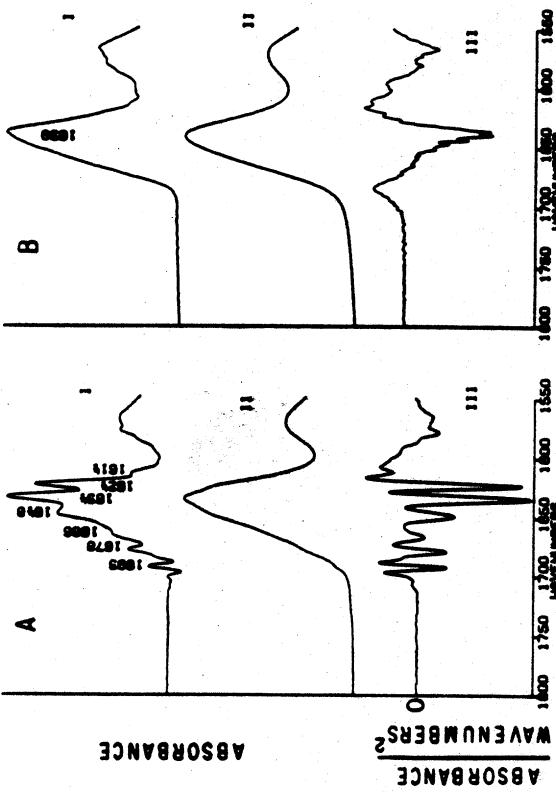


Fig. 7. FTIR spectra of β -lactoglobulin A in various solvents. (A) Native protein in D₂O, 5% w/v; (B) 7% (w/v) in D₂O, 5% w/v, denatured at pH 13; (C) 60% v/v MeOD in D₂O, 0.4% w/v protein; (D) 40% v/v isopropanol-d in D₂O, 0.3%; w/v protein. pH 7 (Purcell and Sung⁴).

under similar conditions (Fig. 1B). It is interesting to note that acidic MeOD has a different effect on the two proteins and that this difference is clearly reflected in the spectra obtained. (The sharp peaks in Fig. 7C observed below 1600 cm⁻¹ are noise due to the weak signal-to-noise ratio of this rather dilute protein solution.) In isopropanol-d solution, as shown in Fig. 7D, β -lactoglobulin A behaves like chymotrypsinogen A. Figure 7 thus shows how FTIR spectra can clearly distinguish between the native form and three denatured forms of the same protein. We know of no other spectroscopic technique that will accomplish this in quite as much detail.

Acknowledgments

The authors are grateful to their colleague James M. Purcell for his collaboration on this project, particularly with respect to the studies on protein denaturation. They also thank Janine N. Brouillette for her technical assistance in preparing samples and obtaining their FTIR spectra.

[14] Protein Secondary Structure Analysis Using Raman Amide I and Amide III Spectra

By ROBERT W. WILLIAMS

Introduction

This chapter describes methods for the estimation of protein secondary structure content—in terms of percentage helix, β -strand, and reverse turn—from a least-squares analysis of Raman amide I and amide III spectra. A statistical analysis of these estimates for proteins with known structures is included to establish the degree of confidence that may be placed on results for other proteins.

The amide I analysis here is a refinement of earlier work,^{1–3} while the amide III analysis is new. Most of these procedures have been automated and Fortran programs for their implementation may be obtained from the author. These programs call well-documented subroutines written by Lawson and Hanson.⁴

¹R. W. Williams, A. K. Dunker, and W. L. Petricolas, *Biophys. J.*, **32**, 232 (1980).

²R. W. Williams and A. K. Dunker, *J. Mol. Biol.*, **152**, 783 (1981).

³R. W. Williams, *J. Mol. Biol.*, **166**, 581 (1983).

⁴C. L. Lawson and R. J. Hanson, "Solving Least Squares Problems," Prentice-Hall, New York, 1974.