Savitzky, A., & Golay, J. J. E. (1964) *Anal. Chem. 36*, 1627-1639.
Schowen, R. L. (1977) in *Isotope Effects on Enzyme-Catalyzed Reactions* (Cleland, W. W., O'Leary, M. H., & Northrop, D. B., Eds.) pp 64-99, University Park Press, Baltimore.
Shiemke, A. K., Loehr, T. M., & Sanders-Loehr, J. (1986) *J. Am. Chem. Soc. 108*, 2437-2443.
van der Meer, R. A., & Duine, J. A. (1986) *Biochem. J. 239*, 789-791.
van der Meer, R. A., Jongejan, J. A., & Duine, J. A. (1987) *FEBS Lett. 221*, 299-304.
van der Meer, R. A., van Wassenaar, P. D., van Brouwershaven, J. H., & Duine, J. A. (1989) *Biochem. Biophys. Res. Commun. 159*, 726-733.
van Iersal, J., van der Meer, R. A., & Duine, J. A. (1986) *Eur. J. Biochem. 161*, 413-419.
Vellieux, F. M. D., & Hol, W. G. J. (1989) *FEBS Lett. 255*, 460-464.
Vellieux, F. M. D., Frank, J. Jzn., Swarte, M. B. A., Groendijk, H., Duine, J. A., Drenth, J., & Hol, W. G. J. (1986) *Eur. J. Biochem. 154*, 383-386.
Vellieux, F. M. D., Huitema, F., Groendijk, H., Kalk, K. H., Frank, J. Jzn., Jongejan, J. A., Duine, J. A., Petratos, K., Drenth, J., & Hol, W. G. H. (1989) *EMBO J. 8*, 2171-2178.
Williamson, P. R., Moog, R. S., Dooley, D. M., & Kagan, H. (1986) *J. Biol. Chem. 261*, 16302-16305.

Comparison of Various Molecular Forms of Bovine Trypsin: Correlation of Infrared Spectra with X-ray Crystal Structures

Steven J. Prestrelski,*,‡,§ D. Michael Byler,∥,⊥ and Michael N. Liebman#

Department of Physiology and Biophysics, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029, Eastern Regional Research Center, U.S. Department of Agriculture, 600 East Mermaid Lane, Philadelphia, Pennsylvania 19118, and Biotechnology Division, AMOCO Technology Corporation, Naperville, Illinois 60566

Received December 28, 1989; Revised Manuscript Received September 6, 1990

ABSTRACT: Fourier-transform infrared spectroscopy is a valuable method for the study of protein conformation in solution primarily because of the sensitivity to conformation of the amide I band (1700–1620 cm^{-1}) which arises from the backbone C=O stretching vibration. Combined with resolution-enhancement techniques such as derivative spectroscopy and self-deconvolution, plus the application of iterative curve-fitting techniques, this method provides a wealth of information concerning protein secondary structure. Further extraction of conformational information from the amide I band is dependent upon discerning the correlations between specific conformational types and component bands in the amide I region. In this paper, we report spectra–structure correlations derived from conformational perturbations in bovine trypsin which arise from autolytic processing, zymogen activation, and active-site inhibition. IR spectra were collected for the single-chain (β-trypsin) and once-cleaved, double-chain (α-trypsin) forms as well as at various times during the course of autolysis and also for zymogen, trypsinogen, and β-trypsin inhibited with diisopropyl fluorophosphate. Spectral differences among the various molecular forms were interpreted in light of previous biochemical studies of autolysis and the known three-dimensional structures of the zymogen, the active enzyme, and the DIP-inhibited form. Our spectroscopic results from these proteins in D_2O imply that certain loop structures may absorb in the region of 1655 cm^{-1}. Previously, amide I′ infrared bands near 1655 cm^{-1} have been interpreted as arising solely from α-helices. These new data suggest caution in interpreting this band. We have also proposed that regions of protein molecules which are known from crystallographic experiments to be disordered absorb in the 1645 cm^{-1} region and that type II β-turns absorb in the region of 1672–1685 cm^{-1}. Our results also corroborate assignment of the low-frequency component of extended strands to bands below 1636 cm^{-1}. Additionally, the results of multiple measurements have allowed us to estimate the variability present in component band areas calculated by curve fitting the resolution-enhanced IR spectra. We estimate that this approach to data analysis and interpretation is sensitive to changes of 0.01 unit or less in the relative integrated intensities of component bands in spectra whose peaks are well resolved.

The ability of biotechnology and protein engineering to produce new and novel proteins has underscored the need for a simple, reliable probe of protein conformation in a variety of conditions and environments. Currently, X-ray crystallography provides the most detailed information concerning positions of individual atoms in the protein structure. However, the length and complexity of the experiments and analysis, as well as the difficulty in crystallizing many proteins, severely limit the use of this tool. Circular dichroism has been widely used as method for studying protein conformation in solution, but its relative insensitivity to certain protein conformations, such as β-sheets, often results in essentially qualitative results (Johnson, 1988).

Fourier-transform infrared spectroscopy has become recognized as a valuable tool for the examination of protein

* Author to whom correspondence should be addressed.
‡ Present address: Amgen Inc., Amgen Center, 1840 Dehavilland Dr., B-2, Thousand Oaks, CA 91320.
§ Mount Sinai School of Medicine of the City University of New York.
∥ Present address: Philadelphia College of Textiles and Science, Department of Chemistry and Physical Sciences, Schoolhouse Lane and Henry Ave., Philadelphia, PA 19144-3497.
⊥ U.S. Department of Agriculture.
AMOCO Technology Corp.

0006-2960/91/0430-133$02.50/0

conformation in aqueous solution. The reliability and wide applicability of Fourier-transform infrared spectroscopy (FTIR)[1] have resulted in greatly expanded use in studies of protein secondary structure. It has been used to study the effects of ligand binding (Arrondo et al., 1988; Trewhella et al., 1989), pH (Casal et al., 1988; Byler & Purcell, 1989a), temperature (Casal et al., 1988; Byler & Purcell, 1989b), pressure (Wong et al., 1989), and solvent denaturation (Purcell & Susi, 1984; Wasacz et al., 1987) on conformation of globular, structural (Payne & Veis, 1988), and membrane-related proteins (Surewicz et al., Surewicz & Mantsch, 1988a; Haris et al., 1989). Most infrared studies of protein conformation focus on absorptions in the amide I (amide I′ in deuterated proteins) region (1700–1620 cm^{-1}) which arise primarily from stretching vibrations of the backbone C=O groups. The frequency of these vibrations has been shown to be sensitive to the molecular geometry and hydrogen-bonding characteristics of the peptide backbone (Miyazawa & Blout, 1961), and specific conformational types (helix, strand, turns, etc.) give rise to discrete bands in the amide I′ region (Byler & Susi, 1986). Component bands in this region are typically broad and lie in close proximity to one another, and the observed amide I band often appears featureless. Thus, mathematical methods termed resolution-enhancement techniques, including derivative spectroscopy (Susi & Byler, 1983) and Fourier self-deconvolution (Byler & Susi, 1986), are necessary to resolve the individual amide I components. These techniques do not increase the instrumental resolution, but mathematically narrow the component bands so that they are more easily visualized. The combination of these methods with the judicious application of curve-fitting techniques can reveal a wealth of information concerning the secondary structure of proteins (Byler & Susi, 1986; Lee & Chapman, 1986; Susi & Byler, 1986; Surewicz & Mantsch, 1988a).

Despite the extensive information which can be extracted by using present methods, the application of FTIR to examining protein conformation is somewhat limited by the lack of correlation between specific backbone folding types and individual component bands in the infrared spectrum. The theoretical studies of model peptide systems performed by Krimm and co-workers [for an excellent review, see Krimm and Bandekar (1986)] provide guidelines for the interpretation of protein IR spectra. However, the inherent complexity, size, and lack of symmetry in globular proteins often prevent direct application of these results to protein spectra. Aided by tools developed for macromolecular structure analysis [e.g., see Levitt and Greer (1977), Kabsch and Sander (1983), and Liebman (1986)], comparisons of IR spectra with high-resolution crystallographically determined protein structures can establish necessary spectra–structure correlations (Byler & Susi, 1986; Arrondo et al., 1988) as well as verify previous assignments which have often been based simply on empirical observations. If additional biochemical information is available, this process can be enhanced.

Trypsin is a well-studied, autolytic serine protease. The process of autolysis in bovine trypsin is well-defined, and the sequence positions of several successive autolytic cleavages have been precisely determined as is outlined in Figure 1 (Schroeder & Shaw, 1968; Smith & Shaw, 1969). Three-dimensional structures of bovine trypsin in several closely related molecular states, including the active enzyme, the zymogen, complexed

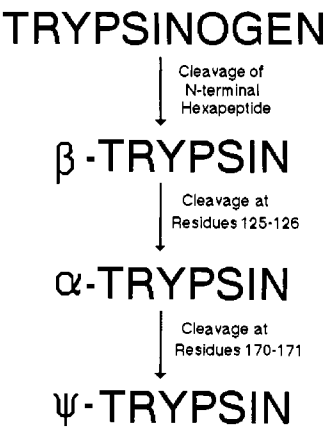

FIGURE 1: Sequential outline of autolytic cleavages in bovine trypsin.

with various small molecule and macromolecular inhibitors, and at several pH values, have been determined by X-ray crystallography (Bernstein et al., 1977). The atomic resolution of a number of these structures is high (1.5–1.8 Å for proteins used in this study). Liebman (1986) has reported an extensive analysis of these structures using tools developed for analysis of secondary and tertiary structure in proteins as well as methods for discerning small conformational differences in closely related molecules such as these. This study reported that the conformational changes which occur upon the transition between the different molecular states of bovine trypsin are localized within well-defined regions of the molecule and, thus, the overall backbone conformation remains unchanged. Thus, this system is well suited for analysis of the IR spectral changes which accompany small conformational changes.

In this paper, we report spectra–structure correlations derived from small, but significant, conformational changes which accompany (1) autolytic cleavage, (2) zymogen activation, and (3) active-site-directed inhibition in bovine trypsin and which result in changes in the amide I′ region of the infrared spectra. Additionally, our studies have enabled us for the first time, to our knowledge, to estimate the variability which results from the collection and analysis of protein FTIR spectra and to comment on the effects that such variability may have on the interpretation of the spectra and their subsequent correlation to protein structure.

Experimental Procedures

Materials. Bovine trypsin (3× crystallized, TRL3) and trypsinogen (1× crystallized, TG) were purchased from Worthington Biochemical Corp., Freehold, NJ. Diisopropyl fluorophosphate was purchased from Sigma Chemical Co., St. Louis, MO. α-Trypsin and β-trypsin were separated from the commercial preparation and purified as described in Schroeder and Shaw (1968). Diisopropylphosphoryl-inhibited β-trypsin was prepared as described in Cunningham (1954) using purified β-trypsin. Trypsinogen was used without further purification. Purified components were stored below 0 °C as lyophilized powders. All other compounds used were reagent grade.

Throughout this report, we will refer to the amino acid sequence positions based on the ordered residue list in β-trypsin, i.e., 1–223 for the 223 residues in β-trypsin. Thus, the seventh residue in trypsinogen, isoleucine, assumes the number 1. The first six residues will be referred to as the N-terminal hexapeptide. The utility of this numbering has been discussed previously (Liebman, 1986).

Spectroscopy. For IR spectroscopy, proteins were prepared as 3.0% (w/v) solutions (1.3 mM) in 20 mM acetate (pD 5.0)

[1] Abbreviations: DIFP, diisopropyl fluorophosphate; DIP, diisopropylphosphoryl; FTIR, Fourier-transform infrared spectroscopy; IR, infrared; PDB, protein data bank; RMS, root mean square; SD, standard deviation.

Table I: Peak Positions (cm^{-1}) and Relative Intensities of Various Molecular States of Trypsin

β-trypsin				α-trypsin				trypsinogen				DIP-β-trypsin			
ν^a	HW[b]	A^c	σ^d	ν	HW	A	σ	ν	HW	A	σ	ν	HW	A	σ
1625	2.5	0.08	0.002	1625	2.6	0.08	0.003	1625	2.4	0.06	0.001	1625	2.9	0.08	0.001
1634	3.1	0.23	0.001	1634	3.5	0.22	0.007	1633	3.3	0.20	0.007	1634	3.4	0.25	0.005
1643	3.9	0.24	0.003	1645	6.2	0.33	0.004	1645	8.0	0.41	0.016	1645	5.5	0.28	0.011
1654	3.4	0.15	0.004	1656	3.2	0.10	0.001	1656	2.6	0.07	0.004	1656	3.0	0.10	0.007
1664	2.9	0.12	<0.001	1664	3.4	0.12	0.001	1665	3.4	0.12	0.003	1665	3.4	0.12	0.001
1674	3.3	0.10	0.001	1674	2.9	0.08	0.003	1674	2.7	0.07	<0.001	1674	2.7	0.07	<0.001
1684	2.9	0.06	<0.001	1684	2.5	0.05	0.002	1683	2.5	0.05	<0.001	1682	2.9	0.06	<0.001
												1690	2.2	0.04	<0.001
1693	2.7	0.02	0.001	1692	3.2	0.02	0.001	1692	3.4	0.02	0.001				

[a] Frequency positions (cm^{-1}) are rounded off to the nearest integer. [b] Half-width at half-height of component bands in the deconvoluted spectra. [c] Mean relative intensity calculated for three independent experiments. [d] Standard deviation for relative intensities calculated for three independent experiments.

or 20 mM imidazole (pD 6.9) buffers made up in D_2O. pD was determined by adding 0.4 to the pH reading (Covington et al., 1968) measured with a Horiba Cardy glass electrode pH meter. The spectrum of each protein was collected at the same pH (pD in this case) as the crystal structure determination. Sufficient $CaCl_2$ (1 M in D_2O) was added to make the final solutions 20 mM in Ca^{2+}. Solutions were placed in IR cells with CaF_2 windows and Teflon spacers with a path length of 75 μm.

IR spectra were collected at ambient temperature using a Nicolet 740 SX FTIR system equipped with a water-cooled Globar source, a Ge-coated KBr beam splitter, and a broad-range mercury/cadmium/telluride detector. All spectra were recorded at a nominal resolution of 2 cm^{-1} by co-adding 4096 double-sided interferograms (0.44 s/scan) which were Fourier-transformed after application of a Happ–Genzel apodization function. The spectrometer and sample chamber were continuously purged with dry nitrogen. Spectral contributions from uncompensated H_2O vapor in the light path and from buffers were subtracted by using programs provided with the Nicolet FTIR software, version 4.3. Factors for water vapor subtraction were determined by subtracting a second derivative spectrum of water vapor from the second derivative spectrum of the sample. The subtraction factor was varied until the region from 1700–1800 cm^{-1} was featureless. Subtraction using second derivative spectra is preferable to the original absorbance spectra because sharp vapor lines, which are often invisible in original spectra, are amplified by differentiation. This results in a more reliable correction for uncompensated water vapor.

For each lyophilized protein sample, except when used in time course studies, three separate solutions were prepared and their spectra collected at different times, often several days apart. Further, each of these spectra was analyzed independently. All spectra were collected as soon as possible after mixing. The time required to load the cell, transfer it to the spectrometer, and purge the sample chamber was approximately 30 min. Collection time for a single spectrum of 4096 co-added interferograms was also 30 min.

Second derivative spectra were calculated analytically as described by Susi and Byler (1983) with the Nicolet software except that the first derivative function was applied twice. Thus, each of the second derivative data points is calculated over five data points rather than three. Fourier self-deconvolutions were also calculated by using the Nicolet software which is based on the method of Kauppinen et al. (1981a). A Lorentzian line shape function and a Bessel apodization function were used. For all proteins studied here, deconvolution parameters of 18 cm^{-1} and 2.8 for the undeconvoluted band half-width and resolution enhancement factor, k, respectively, were found to be optimal. Curve fitting was performed by using the program ABACUS, an iterative Gauss–Newton nonlinear regression program [see Byler and Susi (1986)]. Deconvoluted spectra were fitted with Gaussian band profiles. Gaussian bands, as opposed to Lorentzian or a Gauss–Lorentz combination, were used to fit deconvoluted spectra because the deconvolution procedure, which assumes a Lorentzian band-shape function, removes the Lorentzian contribution to the band shape. The band shape in the deconvoluted spectra, then, takes the shape of the Bessel apodization function which is better approximated by a Gaussian profile (Griffiths et al., 1987). Initial band positions were taken directly from the second-derivative spectra, and no additional bands were added. Initial values for the peak heights and widths were estimated from the deconvoluted spectra. A nonsloping base line was estimated from the spectrum and held constant. For the final fits, the heights, widths, and positions of all bands were varied simultaneously. The relative integrated intensity of each band (i.e., the band area as a fraction of the total amide I′ area) was calculated from the final fitted band heights and widths.

Results

β-Trypsin. Figure 2 shows the amide I′ region of the non-resolution-enhanced spectra of β-trypsin at pD 5.0 as well as its second derivative and Fourier self-deconvoluted spectra. At this pD, autolytic activity is negligible within the time needed for sample preparation and data collection (R. Buono, personal communication). Table I lists the peak positions and areas relative to the total area of the amide I′ region calculated by curve fitting the deconvoluted spectra (see Figure 3A). The predominant feature in the deconvoluted amide I′ region is the intense peak at 1634 cm^{-1}. An additional strong peak appears at 1643 cm^{-1}. Weaker peaks are also present at 1625, 1654, 1663, 1673, 1684, and 1693 cm^{-1}. The second derivative spectrum of β-trypsin resolves two peaks at 1689 and 1695 cm^{-1} in place of the single band at 1693 cm^{-1} in the deconvoluted spectrum. Figure 3A also shows peaks, due to side chain vibrations (Bendit, 1967; Chirgadze et al., 1975), fitted at 1590, 1601, and 1611 cm^{-1}. We routinely include these peaks in the curve-fitting analysis to avoid the approximation otherwise incurred with addition of a sloping base-line parameter. These peaks are not part of the amide I′ band and, therefore, are not used in calculating the relative intensities of its components.

Examination of the ratio of the intensity of the amide II′ band to the intensity of the amide II band showed that for all forms of trypsin studied, hydrogen–deuterium exchange was almost entirely complete. Additional spectra collected at later times showed no differences other than small downward shifts in the frequency (1–2 cm^{-1}) most likely due to exchange of a few more slowly exchanging hydrogens.

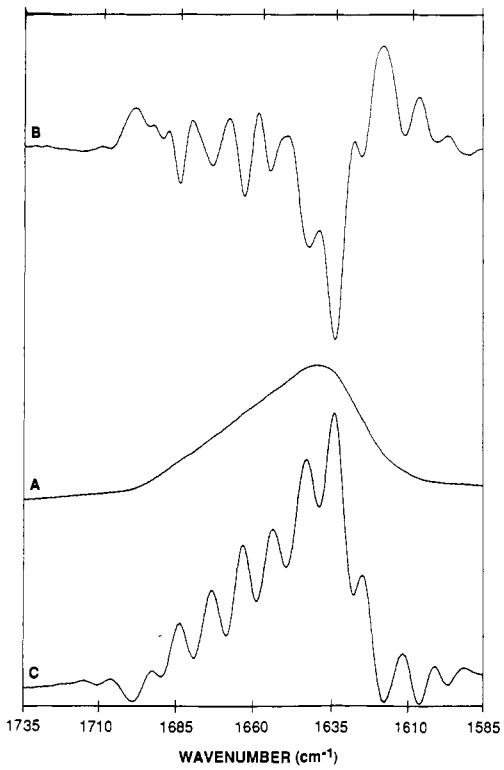

FIGURE 2: Amide I′ region of the infrared spectrum of bovine β-trypsin at pD 5.0, 20 mM Ca^{2+}: (A) non-resolution-enhanced IR absorption spectrum; (B) second derivative spectrum; (C) spectrum after Fourier self-deconvolution.

α-Trypsin. Figure 3B shows the curve-fitted, deconvoluted amide I′ region of α-trypsin at pD 5.0. As with β-trypsin, at this pD autolytic activity is negligible. Table I lists the frequencies of the amide I′ component peaks and their relative areas. The overall shape of the amide I′ region as well as the positions and relative intensities of its components is similar to that of β-trypsin. This is in general agreement with biochemical studies which show α-trypsin to be an enzymatically active species, although with reduced activity (Schroeder & Shaw, 1968). Closer inspection does show some clear distinctions between the spectra of α-trypsin and β-trypsin. The greatest difference occurs at the peak occurring about 1645 cm^{-1}. This peak becomes broader and increases in relative intensity upon conversion from β- to α-trypsin. The relative intensity increases by 0.09. Concomitant with the increase in the intensity of the 1645 cm^{-1} peak are decreases in the intensities of several peaks. The band showing the greatest decrease is 1655 cm^{-1}. Its relative intensity decreases by 0.05. Additionally, the peaks at 1624, 1634, 1673, and 1683 cm^{-1} all decrease slightly, but significantly, in relative intensity.

Some of the second derivative spectra of α-trypsin appeared to have two separate peaks in the 1645 cm^{-1} region. However, the signal-to-noise ratios of these spectra were insufficient to verify this. Therefore, the curve-fitting analysis was carried out considering both possible cases, using one or two peaks in this region. The inclusion of two peaks instead of one did not affect the final results to a significant degree (i.e., the sum of the fractional areas of these two peaks was roughly equal to the fractional of the single peak). Thus, we chose the simpler data set for presentation and interpretation.

Trypsinogen and DIP-β-trypsin. Table I lists the peak positions and relative areas of the amide I′ component bands revealed from curve fitting the deconvoluted spectra of trypsinogen and DIP-β-trypsin (see panels C and D, respectively, of Figure 2). The predominant feature in the deconvoluted amide I′ region of these spectra is the intense peak occurring near 1634 cm^{-1}. An additional strong peak appears around 1644 cm^{-1}. This band has the largest band area in trypsinogen (Table I). Weaker peaks are also present at about 1625, 1655, 1665, 1674, and 1683 cm^{-1} in the spectra of these proteins. Like β-trypsin, the second derivative spectrum of trypsinogen has two peaks at 1689 and 1695 cm^{-1} whereas the deconvoluted spectrum has only a peak at 1693 cm^{-1}. In contrast, both resolution-enhanced spectra of DIP-β-trypsin display only a single peak near 1690 cm^{-1}.

Previously Published Spectra. Spectra of trypsin and trypsinogen published ealier by Byler and Susi (1986, 1988) differ somewhat from the spectra reported here. The enzymes used in those studies were from commercial sources and were not subjected to further purification. In all instances, spectra of these sample were collected a number of hours after dissolution in D_2O at pD ~7. The spectra for trypsin at pD 7 in the presence of Ca^{2+} (Byler & Susi, 1988) most closely resemble the spectra of β-trypsin presented here in Figures 2 and 3A, but the greater intensity in the 1645 cm^{-1} band in the former (due to segments with irregular conformations) suggests that some impurities may be present and/or some autolysis of the sample may have occurred. The spectra of trypsin in which Ca^{2+} was removed by dialysis (Byler & Susi, 1988) can now be understood to be of a sample which has probably undergone extensive autolysis. (Although the enzyme was dialyzed at pD 5 to minimize autolysis, as verified by a check of its activity, the spectra were again collected at pD 7.) In these spectra, the band due to the residues with irregular conformations near 1643 cm^{-1} and that due to β-strands near 1634 cm^{-1} are no longer resolved but appear as a single, broad feature near 1638 cm^{-1}.

Although trypsinogen is less sensitive to autolysis than trypsin, the longer times involved in the earlier work (Byler & Susi, 1986) before the spectra were collected may again have resulted in some degradation of the zymogen. In addition, the older spectra had a lower signal-to-noise ratio than the present data. Thus, less extensive deconvolution was attempted. In turn, fewer amide I′ component bands were resolved, and less reliable and less reproducible curve fits were obtained. In particular, the final value of about 1636 cm^{-1} used in curve fitting in the earlier study was evidently too high and led to distorted band areas for both the β-strands and the irregular components. The frequency observed in the second-derivative spectrum was about 1634 cm^{-1}, in reasonable agreement with the value reported here.

Autolysis of β- and α-Trypsin. Table II lists the peak positions and relative areas in the amide I′ region of β-trypsin and of α-trypsin, revealed through curve fitting its deconvoluted spectra, measured at various times after mixing with the pD 6.9 imidazole buffer. At pD 6.9, trypsin undergoes substantial autolytic activity (R. Buono, personal communication). The spectrum measured at 0.5 h demonstrates that significant autolysis has taken place. The deceased intensity of the peak at 1655 cm^{-1} and the broadening and increased intensity of the 1645 cm^{-1} peak are indicative of substantial α-trypsin being present. Under these conditions, two bands are observed in the second-derivative and deconvoluted spectra at 1691 and 1696 cm^{-1} analogous to those seen only in the second-derivative spectrum at pD 5.0. At 2 h, the spectrum reveals only small changes when compared to the 0.5-h spectrum. At 5 h, the second-derivative and the curve-fitted spectra (not shown) clearly demonstrate the almost complete loss of the peak at 1691 cm^{-1} while the remainder of the spectrum remains relatively unchanged. The spectrum at 20 h reveals numerous

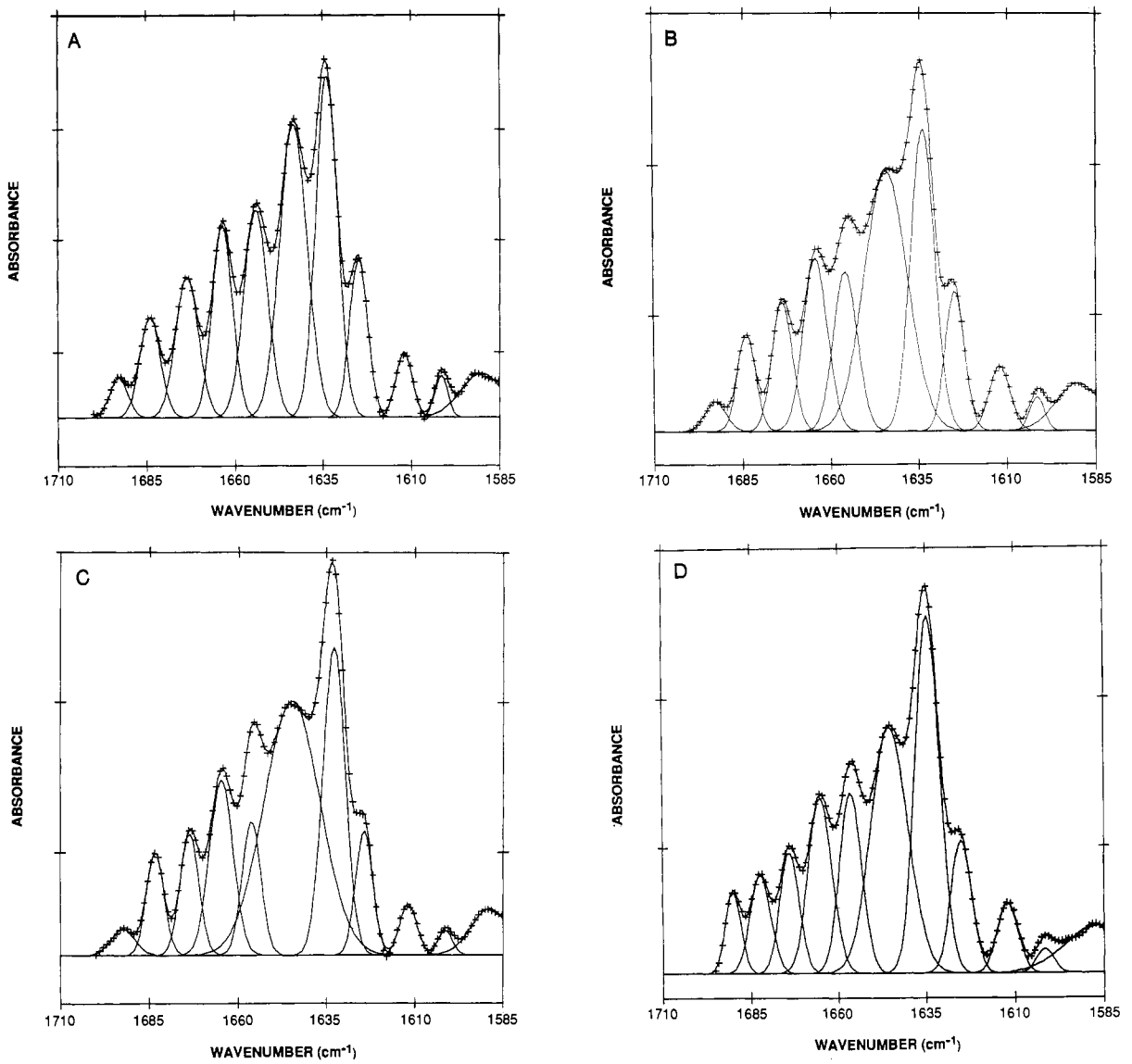

FIGURE 3: Deconvoluted infrared amide I′ band of (A) bovine β-trypsin at pD 5.0, (B) bovine α-trypsin at pD 5.0, (C) bovine trypsinogen at pD 6.9, and (D) DIP-β-trypsin at pD 6.9 (+). Individual Gaussian components and their sum (—). All samples contain 20 mM Ca^{2+}.

Table II: Peak Positions (cm^{-1}) and Relative Intensities of β-Trypsin and α-Trypsin, pD 6.9, at Various Times after Mixing

β-trypsin								α-trypsin					
0.5 h		2 h		5 h		20 h		0.75 h		2.5 h		4 h	
ν	A	ν	A	ν	A	ν	A	ν	A	ν	A	ν	A
1625	0.07	1625	0.07	1625	0.07	1625	0.09	1625	0.07	1625	0.08	1625	0.08
1635	0.27	1634	0.25	1634	0.24	1633	0.15	1635	0.22	1634	0.21	1633	0.14
1647	0.30	1646	0.33	1646	0.33	1643	0.38	1647	0.33	1646	0.35	1643	0.40
1657	0.09	1657	0.09	1656	0.08	1655	0.10	1657	0.10	1656	0.09	1656	0.10
1665	0.12	1665	0.11	1664	0.13	1664	0.13	1665	0.12	1665	0.12	1664	0.13
1674	0.07	1674	0.07	1674	0.07	1674	0.09	1674	0.07	1674	0.08	1674	0.08
1684	0.05	1684	0.05	1684	0.06	1684	0.05	1684	0.05	1684	0.06	1684	0.06
1691	0.02	1691	0.02					1691	0.01				
1696	0.01	1696	0.01	1694	0.01	1694	0.01	1694	0.01	1694	0.02	1694	0.01

small changes which may be indicative of further autolysis and possibly denaturation of the proteins. At this ime, the 1691 cm^{-1} peak has completely disappeared in the deconvoluted as well as the second-derivative spectrum. Table II also lists the peak positions and the relative areas of the component bands of α-trypsin at pD 6.9 measured at various times after mixing. Figure 4 shows the accompanying second derivative spectra. The spectrum measured at 45 min resembles closely the spectrum of α-trypsin at pD 5.0, showing an increase in relative intensity at 1645 cm^{-1} and a decrease in relative intensity at 1657 cm^{-1} compared to the same bands for β-trypsin. Deconvolution of this spectrum also resolves two peaks in the high-frequency region at 1691 and 1694 cm^{-1}. At 2.5 h, the second derivative spectrum reveals a significant loss of intensity of the peak at 1691 cm^{-1} as is also demonstrated in the later stages of autolysis of β-trypsin. This is also evident in the deconvoluted spectrum which shows only a peak at 1694 cm^{-1} although the second derivative still shows a shoulder on the higher frequency side of the 1684 cm^{-1} peak. The peak at 1694 cm^{-1} increases in relative intensity somewhat, but this is probably due to some residual intensity from the 1691 cm^{-1} peak which is no longer resolved. At 4 h, the second derivative

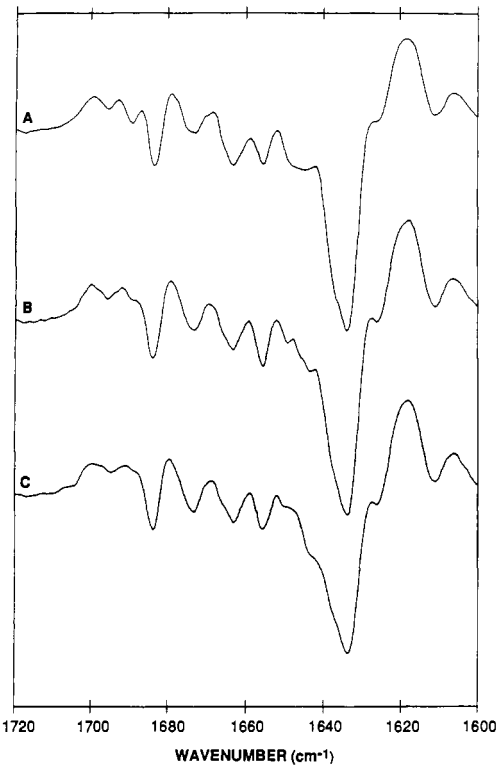

FIGURE 4: Second derivative spectra of the amide I′ region of α-trypsin at pD 6.9, 20 mM Ca^{2+}, at various times after mixing: (A) 45 min; (B) 2.5 h; (C) 4 h. These spectra are normalized so that the peak intensity for the tyrosine band at 1515 cm^{-1} (Bendit, 1967) is constant.

shows the almost complete loss of the peak at 1691 cm^{-1}. In addition to changes in relative intensity, small shifts in frequency occur with time in the spectra of both proteins, but these are invariably toward lower frequency and are most likely isotopic shifts due to further deuteration of a few slowly exchanging peptide hydrogens.

Discussion

Spectral Variability. In studying small differences among spectra of closely related molecules, it is important to gain some measure of the reproducibility or variability among a given set of spectra. For this reason, multiple data sets were collected for each protein sample under carefully controlled conditions as described under Experimental Procedures. These data allow us to assess the inherent variability which is present in the calculated individual band areas (i.e., relative intensities) of protein IR spectra. This evaluation is important in that is allows us to estimate a lower limit for the detection of spectroscopic differences measured by using interferometric infrared techniques and analyzed by using resolution-enhancement and curve-fitting methods. To our knowledge, no previous studies have addressed this question in such detail for IR spectra of proteins.

Various sources of potential variability are present in FTIR measurements of proteins when analyzed with resolution-enhancement and curve-fitting techniques. Variability can be introduced in the sample preparation and data collection processes in many ways. Some of these include small differences in the protein concentration and cell path length, various factors which may lead to instrument instability, the extent of hydrogen–deuterium exchange when using D_2O buffers, and small differences in temperature and pH. Furthermore, small changes in pH and temperature can result in alterations to the buffer spectrum which lead to residual features in the protein spectrum after buffer subtraction. Analysis of measured spectra using mathematical techniques can also introduce variability. Correction of spectra for residual bands due to traces of water vapor in the light path and for buffers is a subjective process. Additionally, noise and uncorrected, residual absorptions of H_2O vapor are very narrow and are amplified relative to the broader protein bands by the resolution-enhancement techniques such as differentiation and self-deconvolution (Mantsch et al., 1989). The process of Fourier self-deconvolution can potentially introduce further variations including nonrandom noise and loss of detail due to smoothing induced by the apodization function. Curve fitting can introduce additional variability because occasionally more than one set of parameters can provide a "fit", particularly if the peaks are not well-defined. The variability in the final calculated band areas, then, is the sum total of all of these contributing factors.

Table I lists the mean and standard deviation for each calculated relative intensity (band area) from three independent experiments. For the component peaks of β-trypsin, α-trypsin, and DIP-β-trypsin, the SD is less than 0.010 for all but the most intense peak in DIP-β-trypsin. Similar results are observed for trypsinogen except for the very broad, intense peak at 1645 cm^{-1} which has an SD of 0.016. This peak, being very broad, is less well-defined in the deconvoluted spectra (Figure 3C). The peaks in the deconvoluted spectra of α-trypsin, β-trypsin and DIP-β-trypsin are quite well-defined. It has been our experience that when peaks are not well-defined, this results in additional variability in the calculated relative intensities because more than one set of parameters may potentially yield good quality fits. Thus, we estimate that if good definition of peaks is achieved in the deconvoluted IR spectra, the calculated relative intensities are sensitive to changes on the order of 0.01 unit (where the total amide I′ band area equals 1.00). Thus, these methods are capable of producing extremely reliable results. It has een estimated that circular dichroism measurements, a widely used method for conformational analysis in proteins, are sensitive to changes on the order of 0.05 unit (R. Buchet, unpublished results).

Stating that these methods are sensitive to changes of 0.01 unit in the relative intensities of amide I′ component bands does necessarily not imply that changes of 0.01 in fractional composition of secondary structures can be detected because the molar absorptivities of the individual conformations have not been established. Progress toward measuring these parameters has been slow due to lack of a suitable experimental systems. However, recent studies (Mantsch et al., 1989; Jackson et al., 1989) and the earlier studies of Byler and Susi (1986) show that the ratios of molar absorptivities among the various conformations are probably close to 1. However, these methods are suitable for observing conformational changes in relative terms, particularly for one protein as it is perturbed by various factors.

The results presented here emphasize the importance of careful data collection and analysis. For studies which examine small spectral differences, it is necessary to collect spectra of sufficiently high signal-to-noise so that smoothing is not necessary. Our experience suggests that for optimal results, the number of scans which are signal-averaged be high enough such that the nonabsorbing region above 1700 cm^{-1} be essentially flat in the second-derivative spectrum (after vapor correction) without additional smoothing. High signal-to-noise spectra are also important in Fourier self-deconvolution as the upper limit of the resolution enhancement factor is determined by the noise level (Kauppinen et al., 1981b). Thus, in order to achieve good definition of component peaks, high signal-

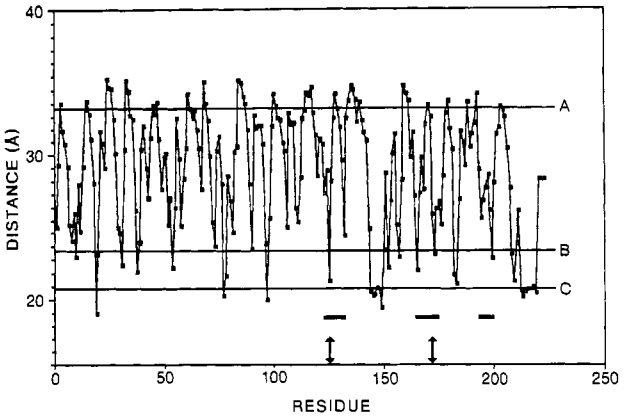

FIGURE 5: Linear distance plot of bovine β-trypsin at pH 5.0. The linear distance value for each residue is calculated by summing the series of distances from its own α-carbon to the α-carbons of the next four residues [see Liebman (1986)]. The horizontal lines in the plot correspond to the values calculated for ideal homopolymers in the (A) β-sheet, (B) 3_{10}-helix, and the (C) α-helix conformations. The three bars under the plot indicate the regions in bovine trypsinogen observed to be disordered in the crystal structure. The two arrows indicate the positions of the autolytic cleavages in bovine trypsin.

to-noise is necessary. Furthermore, careful subtraction of water vapor and buffer is necessary before proceeding to deconvolution because lack of or improper subtraction may lead to spurious bands in the resolution-enhanced spectra (Mantsch et al., 1989). We would also like to point out that, when examining small differences among closely related spectra, a series of spectra should be collected in order to estimate the variability among the spectra and therefore provide confidence that small measured spectral differences arise from conformational differences and not simply variability.

β-Trypsin. Figure 5 displays the linear distance plot of bovine β-trypsin. Analysis of the crystal structure of β-trypsin reveals that the molecule consists primarily of the extended strand conformation (Levitt & Greer, 1977; Marquart et al., 1983; Licbman, 1986). This is also illustrated in the linear distance plot in Figure 5 as segments where the value is greater than 33 Å. Twelve of the extended strands interact in antiparallel β-sheets (Marquart et al., 1983; Liebman, 1986). This agrees well with the intense band in the IR spectrum at 1634 cm^{-1} (Figures 2 and 3A) which has been previously assigned to extended strand structures (Byler & Susi, 1986; Surewicz & Mantsch, 1988). The relative intensity of this band along with the bands occurring at 1625 and 1674 cm^{-1}, which have also been assigned to the extended strand conformation (Byler & Susi, 1986), totals to 0.41. This is in reasonable agreement with estimates of the percentage strand structure in β-trypsin which range from 31% (Marquart et al., 1983) to 56% (Levitt & Greer, 1977). The disagreement between these two methods is due to different criterion for determining a particular conformation. Marquart et al. report only those extended structures which interact to form β-sheets whereas Levitt and Greer report all extended structures. A continuing interest in IR studies of protein conformation lies in determination of which secondary structures contribute to each specific amide I′ component band. In part, the reliability of such an IR technique depends upon having an accepted, self-consistent method for analyzing protein X-ray data to describe the secondary structure of the polypeptide backbone.

Crystal structure analysis shows three α-helices in bovine β-trypsin (Marquart et al., 1983; Liebman, 1986). A single α-helix occurs at residues 144–150, and two distinct but contiguous helices occur at the carboxy terminus. Overall, Liebman (1986) estimates 9% of the residues in β-trypsin adopt the α-helical conformation. These appear in the linear distance plot as stretches of values near 20.5 Å. This agrees well with the results of Levitt and Greer (1977) and Marquart et al. (1983), both of whom estimate 9% α-helical content in β-trypsin. Bands occurring about 1655 cm^{-1} in the IR spectrum have been assigned to the α-helix (Byler & Susi, 1986; Surewicz & Mantsch, 1988a). β-Trypsin has a peak at 1656 cm^{-1} (Figure 3A) which accounts for 0.15 (Table I) of the total amide I′ intensity. This value for α-helix is in general agreement with the crystal structure estimate. Nevertheless, it appears somewhat high when compared with the close agreement demonstrated between the fraction of α-helix determined by an automatic secondary structure analysis algorithm (Levitt & Greer, 1977) and the relative intensity of the band occurring near 1655 cm^{-1} for a variety of globular proteins (Byler & Susi, 1986).

A strong band occurs at 1645 cm^{-1} in the β-trypsin spectra accounting for 0.24 of the amide I′ intensity (Table I). Bands in this region have previously been assigned to "unordered" conformations, or conformations which do not have intrachain hydrogen bonds (Byler & Susi, 1986; Surewicz & Mantsch, 1988a). Instead, the backbone carbonyls may hydrogen-bond to solvent molecules. The structure of such segments of the peptide chain does not conform to any standard type of conformation. The term "unordered", as used in this context, is somewhat of a misnomer, as such conformations do not necessarily lack order in the crystallographic sense. Correlations between specific conformations and this band have not been discerned partly due to this ambiguity in definition. Indeed, X-ray diffraction data display no segment of the β-trypsin which is crystallography disordered. To avoid confusion, in this paper we will use the term "irregular" to refer to conformations which are not classified as helix, extended strand, reverse turns, or loops. The remainder of bands in the β-trypsin spectrum, as in most globular proteins, are tentatively attributed to reverse turns and loops (Byler & Susi, 1986; Surewicz & Mantsch, 1988a).

Autolysis. Figure 1 summarizes the autolysis process in bovine trypsin. The first autolytic cleavage in bovine trypsin, after the one which produces zymogen activation, occurs between residues 125 and 126, when allowing β-trypsin to autodigest. This results in the formation of α-trypsin, a double-chain protein held intact by disulfide bonds. The backbone conformation in this region in the β form of trypsin consists of a loop (often called the "autolysis loop") bounded by stretches of extended strand (Marquart et al., 1983; Liebman, 1986). The greatest difference between β- and α-trypsin (Figure 3A,B) is the loss of intensity in the 1656 cm^{-1} band with an increase in the 1645 cm^{-1} band (Table I). Previously, amide I′ bands around 1655 cm^{-1} for protein in D_2O have been assigned almost exclusively to α-helices (Byler & Susi, 1986; Surewicz & Mantsch, 1988). Examination of the crystal structures shows the α-helices in trypsin are spatially and sequentially distant from this loop (see Figure 5) and are presumably unaffected by this autolytic cleavage. Thus, the decrease in the intensity of the 1655 cm^{-1} band after autolysis appears to correlate with the cleavage of the loop rather than from any change in the amount of helical structure. This observation suggests that in proteins certain loops may also absorb around 1655 cm^{-1}. Further evidence supporting this interpretation lies in the relative intensity of this band in the two proteins. In β-trypsin, the relative intensity of the 1656 cm^{-1} band is 0.15 and 0.10 in α-trypsin (Table I). As stated above, X-ray data indicate that 9% of trypsin is in the α-helical conformation. Additional evidence for such an interpretation

is found in Surewicz and Mantsch (1988b,c) and Surewicz et al. (1988). Based on an unexpected band around 1655 cm^{-1} in several small peptides and proteins which were determined by circular dichroism to contain no helices, these investigators have suggested that this peak arises from an unexchanged irregular conformation (Susi, 1969) or an "atypical nonperiodic" conformation. Our results would suggest the latter because the autolysis loop lies at the surface where it is susceptible to autolytic cleavage and should therefore be available for hydrogen exchange. Thus, these results suggest caution in interpreting bands near 1655 cm^{-1} as resulting exclusively from α-helices.

The increased intensity of the band near 1645 cm^{-1} upon conversion from β- to α-trypsin (Table I) is consistent with its assignment to "irregular" structures (Byler & Susi, 1986; Surewicz & Mantsch, 1988). Local unfolding is expected at the part of the protein molecule close to the site of cleavage of a peptide bond. This assumption is further corroborated by crystallographic data from trypsin which have been inhibited with diisopropyl fluorophosphate (Chambers & Stroud, 1979). In this determination, crystals of this compound were observed to contain an approximately 50–50 distribution of β- and α-trypsin. The electron density in the region of the autolysis loop is weak, indicating disordering in the cleaved, two-chain component (i.e., α-trypsin). Such an interpretation is also consistent with previous studies which indicate broadening and increased fractional intensity of bands near 1645 cm^{-1} with an increase in irregular or disordered structures (Purcell & Susi, 1984; Byler & Susi, 1986; Surewicz et al., 1987). Such broadening is most likely due to a greater population of slightly different hydrogen-bonding patterns between solvent molecules and polypeptide fragments lacking well-defined, repetitive structures (Surewicz & Mantsch, 1988a).

Other spectroscopic changes which occur upon conversion from β- to α-trypsin are small decreases in intensity in the 1625, 1634, 1673, and 1684 cm^{-1} bands (Figure 3A,B). The 1625 and 1634 cm^{-1} bands are assigned to extended strands. Byler and Susi (1986) assigned bands at 1675 cm^{-1} as the high-frequency component of β-strands, but the precise position of this component still remains unclear although the 1673 cm^{-1} band is a likely choice. The loss in intensity of the bands assigned to β-strands is probably due to disordering or disruption in hydrogen bonding of the β-strands which bound this loop. We cannot assign the 1684 cm^{-1} band to a specific conformation with presently available data.

A second autolytic cleavage in bovine trypsin occurs between residues 170 and 171 (see Figure 1) when allowing α-trypsin to autodigest (Smith & Shaw, 1969). The resulting twice-cleaved product has been designated ψ-trypsin. Unlike the conformation in the region of the first autolytic cleavage, the conformation in this region is complex and contains several types of regular secondary structure. It consists of a short stretch of extended strand bounded by a 3_{10}-loop or type III β-turn on the N-terminal side and two consecutive standard type II β-turns on the C-terminal side (Chambers & Stroud, 1979). Spectroscopically, the most significant changes upon conversion of α-trypsin to ψ-trypsin are loss of the peak at 1691 cm^{-1} and loss of intensity at 1633 cm^{-1} and an increase in relative intensity at 1644 cm^{-1} (see Figure 4, Table III). After 4 h of incubation, ψ-trypsin becomes the predominant species (Smith & Shaw, 1969). Loss of intensity in the 1634 cm^{-1} band and gain in intensity in the 1644 cm^{-1} band can be interpreted as a loss of extended-strand structure either from the strand in the region of the cleavage or from some combination of the two. Because of the complexity of the folding

Table III: Amide I′ Spectra–Structure Assignments for Bovine Trypsin

band frequency (cm^{-1})	conformation
1625	extended strand[a,b]
1635	extended strand[a,b]
1645	irregular,[a,b] disordered[a]
1655	α-helix,[b] loops[a]
1664	turns[b]
1674	extended strand[b] possibly type II β-turns
1683	turns,[b] possibly type II β-turn[a]
1689	turns[b]
1695	turns,[b] possibly carboxyl C=O[c]

[a] Proposed or corroborated from this study. [b] Assigned in previous investigations; see Surewicz and Mantsch (1988) and references cited therein. [c] See Casal et al. (1989).

in this region, the origin of the 1691 cm^{-1} band remains unclear. In studies of model peptide systems, Krimm (1986) has suggested that certain types of β-turns can absorb in this high-frequency region. While this is consistent with the presence of turn structures in this portion of the molecule, the lack of more definite information concerning structural changes in this complexly folded region prohibits unequivocal assignment of the 1691 cm^{-1} band to a specific conformation.

Another result which merits discussion is the observation that the spectrum of β-trypsin after 20 h of incubation and the spectrum of α-trypsin after 4 h of incubation (see Table II) are essentially identical, except for small differences in the position of some of the bands. This is what one might expect as it is known that after conversion to α-trypsin, the solution initially containing β-trypsin should then convert to ψ-trypsin. This result illustrates the utility of FTIR spectroscopy in monitoring biochemical processes which involve relatively small changes in protein conformation.

Trypsinogen to Trypsin Transition. Conversion of trypsinogen to β-trypsin occurs as a result of an autolytic cleavage of the protein chain which releases the N-terminal hexapeptide (see Figure 1). This cleavage results in a series of conformational changes which have been described as a disorder-to-order transition (Huber & Bode, 1978). Three regions of the trypsinogen molecule, collectively termed the activation domain, are known from crystallographic experiments to be disordered (Felhammer et al., 1977; Walter et al., 1982). These regions are identified in Figure 5. These residues become ordered, although not necessarily adopting regular secondary structure, to form the active enzyme, β-trypsin. The 3 regions which make up the activation domain, residues 122–132, 164–175, and 193–200, account for 31 residues. In addition, first six residues in trypsinogen, the N-terminal hexapeptide, are observed to be disordered in the trypsinogen molecule (Fehlhammer et al., 1977). Thus, the total number of residues disordered in trypsinogen is 37. The term disordered as used here refers to crystallographic disorder which means that in the crystallographic experiments no significant electron density is observed for these regions. This is due to conformational flexibility in these regions or the presence of multiple stable conformations (Walter et al., 1982). The term "disordered", thus, contrasts with the term "irregular". We shall use the term "disordered" only to refer to regions which are crystallographically disordered. The remainder of the trypsinogen molecule, which accounts for 84% of the residues, has almost identical conformation when compared to β-trypsin using a difference linear distance technique (Liebman, 1986), and the RMS deviation is only 0.2 Å upon structural superposition of these regions (Huber & Bode, 1978).

The greatest difference between the amide I region of the IR spectra of trypsinogen and β-trypsin occurs in the band near

1645 cm^{-1}. This band is much broader (Figure 2C) and greater in relative intensity in the trypsinogen spectra (Table I). Its relative intensity is 0.41 versus 0.24 for β-trypsin, an increase of 0.17 upon conversion to the active enzyme. As noted in our discussion of the previous two sections, such amide I′ bands have often been assigned to unordered or irregular conformations in proteins. Such an assignment is consistent with the changes in bandwidth and intensity observed here upon zymogen activation. Ordering of the 31 residues of the activation domain and the loss of 6 disordered residues of the N-terminal hexapeptide, a total of 16% of the residues, correlate very well with a 0.17 decrease in relative intensity of this band. In this case, our results imply that assignment of bands around 1645 cm^{-1} to "unordered" or "irregular" conformations is also consistent with the presence of "disorder" in a crystallographic sense.

To reiterate our previous argument, the presence of a peak near 1645 cm^{-1} in the spectrum of β-trypsin, for which no regions of the crystal exhibit "disorder", suggests that other "irregular" conformations also absorb in this part of the spectrum. Neither "disordered" structures (i.e., those which are not observed crystallographically due to increased motion or the existence of multiple stable conformations) nor "irregular" structures (i.e., those which are observed crystallographically but do not adopt standard, well-defined, intramolecularly hydrogen-bonded conformations) are expected to result in significant transition dipole coupling interactions (Krimm & Bandekar, 1986). Thus, both should absorb in the same frequency range near 1645 cm^{-1}. At present, one cannot distinguish between these two structure classes with IR experiments alone.

The large decrease in relative intensity in the 1645 cm^{-1} band occurs concomitant with increases in the relative intensity of several bands upon conversion from trypsinogen to β-trypsin. This is the only amide I′ component band which shows a decrease in relative intensity upon this conversion. Bands near 1625, 1634, 1654, 1674, and 1684 cm^{-1} (Figure 3C,A) show increases in relative intensity (Table I). These increases in intensity are presumably due to formation of the ordered conformations observed crystallographically in the activation domain of β-trypsin.

Three of these bands which increase in relative intensity upon conversion from trypsinogen to β-trypsin are those near 1625, 1634, and 1675 cm^{-1} which are associated with extended-strand structures (Byler & Susi, 1986). These results are consistent with such an assignment as the X-ray structures of trypsinogen and β-trypsin indicate an increase in extended-strand structures upon zymogen activation. For trypsinogen, estimates of β-strand content from X-ray data range from 29 (Bernstein et al., 1977) to 49%[2] (Levitt & Greer, 1977), compared to a range of 31 (Marquart et al., 1983) to 56% (Levitt & Greer, 1977) in β-trypsin. The sum of the fractional intensities for the three bands assigned to β-strands (33% in trypsinogen; 41% in β-trypsin) increases by 0.08 upon this conversion (Table I). Although the individual estimates of β-strand content vary widely for both proteins, both studies report an increase in β-structure content upon activation of trypsinogen.

One of the newly ordered regions is the autolysis loop of β-trypsin. The autolysis experiments reported here suggest that this loop absorbs near 1655 cm^{-1}. The observed increase in the relative intensity of the band around 1655 cm^{-1} and decrease of the band at 1645 cm^{-1} with the trypsinogen–trypsin transition (Table I) are consistent with this assignment. In trypsinogen, where this segment of the peptide chain is disordered, the relative intensity of the band near 1655 cm^{-1} is of significantly lower relative intensity. While bands in the IR spectrum around 1655 cm^{-1} have been previously assigned solely to α-helices (Byler & Susi, 1986; Surewicz & Mantsch, 1988a), the crystal structures of trypsinogen and trypsin show no conformational differences in the helical regions (Huber & Bode, 1978; Liebman, 1986), providing further evidence for assignment of the loop to the 1655 cm^{-1} band.

The two other flexible loops in the activation domain of trypsinogen also become ordered in trypsin, one forming an extended strand ending in a type II β-turn and the other forming a loop which also contains a type II β-turn. Amide I bands near 1625 and 1634 cm^{-1} have been assigned to low-frequency vibrations of extended strands (Byler & Susi, 1986; Surewicz & Mantsch, 1989). Observed increases in the relative intensities of the bands found near these frequencies after conversion of trypsinogen to β-trypsin are consistent with this assignment. A high-frequency vibration near 1675 cm^{-1} is also expected for extended-strand structures (Byler & Susi, 1986), but because of overlap with bands thought to arise from reverse turns, unequivocal assignment of a band has yet to be determined for this mode (Surewicz & Mantsch, 1988a). Increases in relative intensity are observed at bands near 1674 and 1683 cm^{-1}. One of these presumably arises from the high-frequency extended-strand mode and the other from the β-turns formed. At present, specific classes of reverse turns have not been assigned to individual bands in the amide I region. However, these data imply that type II β-turns absorb either near 1675 cm^{-1} or near 1683 cm^{-1}. This suggestion is consistent with theoretical studies of model peptide reverse turn structures performed by Krimm and Bandekar (1986) which predict that type II β-turns have absorptions in the 1680 cm^{-1} region.

Inhibition with DIFP. In comparing crystal structures, Liebman (1986) has reported that the two regions of greatest conformational perturbation in β-trypsin upon inhibition with the active-site-directed protease inhibitor DIFP involve residues 122–130 and residues 190–197. Significant spectral changes also accompany this inhibition (Figure 3A,D, Table I). The spectrum of the inhibited form displays increases in the relative intensity of bands at 1644 and 1690 cm^{-1} concomitant with decreases in intensity of bands at 1655 and 1674 cm^{-1}. Additionally, a band near 1695 cm^{-1} is present in the spectrum of β-trypsin but is not present in DIP-β-trypsin.

One of the structurally perturbed regions in DIP-β-trypsin, residues 122–130, is part of the previously described activation domain in trypsinogen (Fehlhammer et al., 1977) and is also the location of the "autolysis loop" in the native β-trypsin (Schroeder & Shaw, 1968). The results of a comparison of the spectra of β-trypsin, trypsinogen, and α-trypsin, as discussed in the previous sections, suggest that this loop in the β-trypsin contributes to the relative intensity of the band at 1655 cm^{-1}. The results observed for DIP-β-trypsin further corroborate this assignment as the relative intensity of the band near 1655 cm^{-1} is significantly lower compared with β-trypsin. Again, no perturbation is observed crystallographically in the helical regions of DIP-β-trypsin (Huber & Bode, 1978; Liebman, 1986). This corroborates the assignment of this loop to the 1655 cm^{-1} region proposed in the autolysis and zymogen activation studies discussed above.

On the basis of the observed intensity changes and structural alterations, we tentatively conclude the autolysis loop in β-

[2] This value was calculated by taking the β-strands determined by Levitt and Greer's analysis of β-trypsin (Levitt & Greer, 1977) and then deleting those segments which in trypsinogen are part of the diordered activation domain.

trypsin contributes to absorptions in the 1655 cm^{-1} region by analogy to two other systems, zymogen activation and autolysis. However, in the absence of specific assignments to the other perturbed chain segments, both in β-trypsin and in DIP-β-trypsin, the complexity of the spectral changes which are observed upon inhibition with DIFP does not allow unambiguous assignment of such conformations to specific spectral bands. Further, some recently published results (Casal et al., 1988; Byler & Purcell, 1989a) have suggested that bands near 1695 cm^{-1} in protein infrared spectra are not amide I' vibrations but represent vibrational transitions of unionized carboxyl groups. Such a possibility further complicates the assignment process in these proteins. On this basis, it might be argued that the absence of this band in DIP-β-trypsin at pD 7 may be due to the increase of the pD relative to that of the spectrum of β-trypsin at pD 5.0. However, a spectrum of β-trypsin at the higher pD (not shown) does not result in loss of this band. Nor does the spectrum of DIP-β-trypsin collected at pD 5.0 (not shown) indicate the presence of a band near 1695 cm^{-1}. Thus, the assignment of this band in the bovine trypsin spectrum remains uncertain.

Table III summarizes the spectra–structure assignments for the molecules examined in this study. These assignments may prove useful for future infrared studies of protein conformation. However, future studies will be necessary to verify the general applicability of these assignments to protein molecules.

CONCLUSION

We have demonstrated that FTIR spectroscopy, combined with resolution-enhancement and curve-fitting techniques, is a very sensitive probe of secondary structure in proteins. One can examine the often minute effects of ligands and environment factors on conformation. To further increase the utility of FTIR for examining protein conformation, it is necessary to discern the relationships between the presence of specific conformational types and the pattern and intensity of bands in the conformation-sensitive amide regions. In the absence of rigorous theoretical methods such as normal mode analysis, which are presently limited by lack of computational power to the study of relatively short peptides, other methods of discerning these relationships are needed. We have demonstrated that comparison of spectroscopic changes with crystallographically determined structures, along with additional biochemical information, can provide a means of ascertaining necessary spectra–structure correlations. In particular, we have shown that certain loop structures may contribute to amide I' bands around 1655 cm^{-1}, previously assigned solely to α-helices. In addition, our studes suggest that, in D_2O, crystallographically "disordered" regions absorb near 1645 cm^{-1} along with "irregular" segments. We have also provided further evidence for previous assignments of bands to extended strand. Future progress toward spectra–structure assignments may come from application of methods exemplified here to other similar systems.

ACKNOWLEDGMENTS

We are grateful to Chad Haggerty, USDA–ARS–Eastern Regional Research Center, Philadelphia, PA, for his help in developing a version of ABACUS which works well with our IR data and to Dr. J. B. A. Ross, Department of Biochemistry, Mt. Sinai School of Medicine, City University of New York, for his helpful comments concerning the format of the manuscript. We also acknowledge Dr. Richard Schultz and the members of the Loyola University Stritch School of Medicine, Department of Biochemistry and Biophysics, for extensive assistance in purifying and preparing many of the trypsin derivatives. Finally, we thank R. Buono, Department of Pharmacology, Mt. Sinai School of Medicine, City University of New York, for providing us with his unpublished results on the autolysis of trypsin.

Registry No. Trypsin, 9002-07-7; trypsinogen, 9002-08-8.

REFERENCES

Arrondo, J. L., Young, N., & Mantsch, H. H. (1988) *Biochim. Biophys. Acta 952*, 261–268.
Bendit, E. G. (1967) *Biopolymers 5*, 525–533.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Yasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Byler, D. M., & Susi, H. (1986) *Biopolymers 25*, 469–487.
Byler, D. M., & Susi, H. (1988) *J. Indust. Microbiol. 3*, 73–88.
Byler, D. M., & Purcell, J. M. (1989a) *Spectrosc. Biol. Mol., Proc. 3rd Eur. Conf.*, 21–24.
Byler, D. M., & Purcell, J. M. (1989b) *Proc. SPIE—Int. Soc. Opt. Eng. 1145*, 415–417.
Casal, H. L., Kohler, U., & Mantsch, H. H. (1988) *Biochim. Biophys. Acta 957*, 11–20.
Chambers, J. L., & Stroud, R. M. (1979) *Acta Crystallogr. B35*, 1861–1869.
Chirgadze, Y. N., Fedorov, O. V., & Trushina, N. P. (1975) *Biopolymers 14*, 679–694.
Covington, A. K., Paabo, M., Robinson, R. A., & Bates, R. G. (1968) *Anal. Chem. 40*, 700–706.
Cunningham, L. W., Jr. (1954) *J. Biol. Chem. 13*, 211–218.
Fehlhammer, H., Bode, W., & Huber, H. (1977) *J. Mol. Biol. 111*, 415–438.
Griffiths, P. R., Pierce, J. A., & Hongjin, G. (1987) in *Computer-Enhanced Spectroscopy* (Meazelear, H. L. C., & Elsenhour, T. L., Eds.) pp 29–54, Plenum Press, New York.
Haris, P. I., Coke, M., & Chapman, D. (1989) *Biochim. Biophys. Acta 995*, 160–167.
Huber, R., & Bode, W. (1978) *Acc. Chem. Res. 11*, 114–122.
Jackson, M., Haris, P. I., Chapman, D. (1989) *Biochim. Biophys. Acta 998*, 75–79.
Johnson, W. C. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 145–167.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Kauppinen, J. K., Moffatt, D. J., Mantsch, H. H., & Cameron, D. C. (1981a) *Appl. Spectrosc. 35*, 271–277.
Kauppinen, J. K., Moffatt, D. J., Cameron, D. C., & Mantsch, H. H. (1981b) *Appl. Opt. 20*, 1866–1879.
Krimm, S., & Bandekar, J. (1986) *Adv. Protein Chem. 38*, 181–364.
Lee, D. C., & Chapman, D. (1986) *Biosci. Rep. 6*, 235–255.
Levitt, M., & Greer, J. (1977) *J. Mol. Biol. 114*, 181–239.
Liebman, M. N. (1986) *Enzyme 36*, 115–140.
Mantsch, H. H., Surewicz, W. K., Muga, A., Moffatt, D. J., & Casal, H. L. (1989) *Proc. SPIE—Int. Soc. Opt. Eng. 1145*, 580–581.
Marquart, M., Walter, J., Diesenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr. B39*, 480–490.
Miyazawa, T., & Blout, E. R. (1961) *J. Am. Chem. Soc. 83*, 712–719.
Payne, K. J., & Veis, A. (1988) *Biopolymers 27*, 1749–1760.
Purcell, J. M., & Susi, H. (1984) *J. Biochem. Biophys. Methods 9*, 523–527.
Schroeder, D. D., & Shaw, E. (1968) *J. Biol. Chem. 243*, 2943–2949.
Smith, R. L., & Shaw, E. (1969) *J. Biol. Chem. 244*, 4707–4712.

Surewicz, W. K., & Mantsch, H. H. (1988a) *Biochim. Biophys. Acta 952*, 115-130.
Surewicz, W. K., & Mantsch, H. H. (1988b) *Biochem. Biophys. Res. Commun. 150*, 245-251.
Surewicz, W. K., & Mantsch, H. H. (1988c) *J. Am. Chem. Soc. 110*, 4412-4414.
Surewicz, W. K., Moscarello, M. A., & Mantsch, H. H. (1987) *Biochemistry 26*, 3881-3886.
Surewicz, W. K., Stepanik, T. M., Szabo, A. G., & Mantsch, H. H. (1988) *J. Biol. Chem. 263*, 786-790.
Susi, H. (1969) in *Structure and Stability of Biological Macromolecules* (Timasheff, S. N., & Fasman, G. D., Eds.) pp 576-663, Marcel Dekker, New York.
Susi, H., & Byler, D. M. (1983) *Biochem. Biophys. Res. Commun. 115*, 391-397.
Susi, H., & Byler, D. M. (1986) *Methods Enzymol. 130*, 290-311.
Trewhella, J., Liddle, W. K., Heidorn, D. B., & Strynadka, N. (1989) *Biochemistry 28*, 1294-1301.
Walter, J., Steigmann, W., Singh, T. P., Bartunik, H., Bode, W., & Huber, H. (1982) *Acta Crystallogr. B38*, 1462-1472.
Wasacz, F. M., Olinger, J. M., & Jakobsen, R. J. (1987) *Biochemistry 26*, 1464-1470.
Wong, P. T. T.; Saint Girons, I., Guillou, Y., Cohen, G. N., Barzu, O., & Mantsch, H. H. (1989) *Biochim. Biophys. Acta 996*, 260-262.

Circular Dichroism Study on the Conformational Stability of the Dimerization Domain of Transcription Factor LFB1

Raffaele De Francesco,*,‡,§,∥ Annalisa Pastore,§ Giuseppe Vecchio,⊥ and Riccardo Cortese§,∥

European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, D-6900 Heidelberg, Federal Republic of Germany, and Istituto di Chimica degli Ormoni, Consiglio Nazionale delle Ricerche, via Maria Bianco 9, I-20131 Milano, Italy

Received June 29, 1990; Revised Manuscript Received September 6, 1990

ABSTRACT: LFB1, a dimeric DNA binding protein, is a major determinant of hepatocyte-specific transcription. The thermal and chemical equilibrium unfolding of a 32-residue α-helical peptide comprising its dimerization domain (B1-Dim) was monitored by circular dichroism spectroscopy. The conformational stability of this peptide is shown to be concentration dependent, and the unfolding reaction is described as a two-state transition between folded dimers and unfolded monomers. The thermodynamic parameters associated with the unfolding reaction were determined under the two-state assumption by the van't Hoff procedure. The enthalpy of unfolding increases linearly with temperature, and the corresponding value of ΔC_p, the difference in heat capacity between the unfolded and the folded forms of the peptide, is estimated to be ca. 0.7 kcal mol^{-1} K^{-1}. The dimeric folded structure of the peptide is stabilized, at 25 °C, by a ΔG of about 11.5 kcal mol^{-1}, which is equivalent to a dimerization constant greater than 10^8 mol^{-1}. These results indicate that the dimerization domain of LFB1 can fold and dimerize independently of the rest of the protein, with a thermodynamic stability comparable to that of a small globular protein.

LFB1, also named hepatocyte nuclear factor 1 (Courtois et al., 1987), is a sequence-specific DNA binding protein that acts as a trans-activator of RNA polymerase II transcription of several liver-specific genes in higher eukaryotes (Monaci et al., 1988; Lichtsteiner & Schibler, 1989). Previous studies (Frain et al., 1989; Nicosia et al., 1990) showed that LFB1 contains a homeobox-like motif and binds to DNA as a dimer. Furthermore, in solution, the dimer exists in equilibrium with the monomer independently of DNA binding (Nicosia et al., 1990).

Many DNA binding proteins that trans-activate transcription in eukaryotes have been shown to form either homodimers or heterodimers with related factors: this feature is of great biological significance with respect to the question of how the activity of these proteins is in its turn regulated and modulated (Johnson & McKnight, 1989). The rôle of dimerization in determining the biological activity and/or specificity of transcriptional activators has been addressed in a recent review (Jones, 1990).

We characterized extensively the DNA binding and dimerization properties of LFB1 by site-directed mutagenesis (Nicosia et al., 1990). The portion of the protein sufficient to confer sequence-specific DNA binding activity spans amino acids 1-281. The DNA binding domain, as defined by mutational analysis, is in turn comprised of three distinct regions, the integrity of which is a prerequisite for functionality. One of these regions, corresponding to the N-terminal portion of the protein (the "A-domain"), turns out to be necessary and sufficient to guide the dimerization of LFB1. On the basis of sequence analysis and homology to a part of the myosin rod, this region was predicted to form a dimer of α-helices with the hydrophobic amino acids of the common interface packed in a coiled-coil fashion (Cohen & Parry, 1990). Such an arrangement of amphipathic α-helices provides the dimer interface for many DNA binding proteins containing a regular repeat of leucines or "leucine zipper" (Landschulz et al., 1988). However, the N-terminal region of LFB1 shows no sequence homology to the leucine zipper motif.

* To whom correspondence should be addressed.
‡ Recipient of an EMBO long-term fellowship.
§ European Molecular Biology Laboratory.
∥ Present address: IRBM, Istituto di Ricerche di Biologia Molecolare, 00040 Pomezia, Italy.
⊥ Istituto di Chimica degli Ormoni.

0006-2960/91/0430-143$02.50/0