AP

REVIEW

Spectroscopic Methods for Analysis of Protein Secondary Structure

John T. Pelton[1] and Larry R. McLean
Hoechst Marion Roussel, Route 202-206, Bridgewater,
New Jersey 08807-0800

Several methods for determination of the secondary structure of proteins by spectroscopic measurements are reviewed. Circular dichroism (CD) spectroscopy provides rapid determinations of protein secondary structure with dilute solutions and a way to rapidly assess conformational changes resulting from addition of ligands. Both CD and Raman spectroscopies are particularly useful for measurements over a range of temperatures. Infrared (IR) and Raman spectroscopy require only small volumes of protein solution. The frequencies of amide bands are analyzed to determine the distribution of secondary structures in proteins. NMR chemical shifts may also be used to determine the positions of secondary structure within the primary sequence of a protein. However, the chemical shifts must first be assigned to particular residues, making the technique considerably slower than the optical methods. These data, together with sophisticated molecular modeling techniques, allow for refinement of protein structural models as well as rapid assessment of conformational changes resulting from ligand binding or macromolecular interactions. A selected number of examples are given to illustrate the power of the techniques in applications of biological interest.   © 2000 Academic Press

Since the early X-ray studies of Kendrew on myoglobin revealed the folding of the polypeptide chain, biochemists have worked to relate the amino acid sequence of a protein to its three-dimensional structure. Despite great effort, this goal remains elusive. However, the need to predict the conformation of proteins is becoming ever more acute as the flood of new sequences arising from the human genome project provides an array of possible new targets for biomedical research. Yet, we know the detailed structure of less than 3% of the 100,000 protein genes whose primary structure (amino acid sequence) have been determined. Many of these uncharacterized genes have sufficient sequence similarity to known proteins to permit a relatively accurate prediction of protein structure by homology modeling. Other sequences have less than 25% identity to any known protein and may adopt novel folds. These folds comprise multiple domains of defined secondary structure that are expected to be similar to those of known proteins. As a result, experimental determination of the secondary structure of a protein may provide important clues to understanding its function.

A set of secondary structural elements defines a protein motif. Experimental verification of a predicted folding motif may be gained by measurements of protein secondary structural elements of which the motif is composed. In many cases the fraction of peptide bonds in α-helical, β-pleated sheet, and aperiodic conformations may be estimated from highly sensitive optical measurements, such as CD[2] (circular dichroism), IR (infrared) and Raman spectroscopies. In favorable cases, the spectral effects of conformational changes that accompany substrate and inhibitor binding, subunit assembly, and other biological phenomena may

---

[1] To whom correspondence should be addressed at Hoechst Marion Roussel, Room N-1209, Route 202-206, Bridgewater, NJ 08807-0800. Fax: (908) 231-3576. E-mail: tom.pelton@hmrag.com.

[2] Abbreviations used: GdnHCl, guanidine hydrochloride; CD, circular dichroism; IR, infrared; NMR, nuclear magnetic resonance; chaps, 3-[(3-cholamidopropyl)dimethyl ammonio]propanesulfonate; PEM, piezoelastic modulator; PM, photomultiplier; SVD, singular value decomposition; CCA, convex constraint analysis; ATR, attenuated total reflectance; FFT, fast Fourier transform; FSD, Fourier self-deconvolution; FWHH, full bandwidth at half height; CLS, classical least squares; PLS, partial least squares; CSI, chemical shift index; VAMP, vesicle-associated membrane protein; SNAP-25, synaptosome-associated protein-25.

also be observed. Such measurements provide valuable information for site-directed mutagenesis and confirmation of folding in expressed proteins. In addition, spectroscopic measurements of secondary structures are a valuable tool for assessing protein aggregation and stability. For a number of important problems in molecular biology, such as protein folding, protein-protein interactions, and protein-nucleic acid interactions, quantitative measurements of secondary structure provide significant insight into structural features critical to biological function.

This review is designed to provide the reader with an overview of the several experimental techniques currently available to estimate protein secondary structure and assess changes in structure as a result of internal or external factors. Although mass spectrometry and fluorescence spectroscopy have been used to monitor conformational changes, the focus of this review is on experimental methods that have generally been employed to assign secondary structures to proteins. Thus, the discussion is limited to CD, infrared, and Raman spectroscopies as well as NMR (nuclear magnetic resonance) chemical shifts and does not include molecular modeling. The "low-resolution" structures resulting from these techniques clearly do not provide the detailed structural information that is obtained from either X-ray crystallography or high-resolution NMR. Nevertheless, they can help to bridge the gap between amino acid sequence and function by providing clues to the structure a protein may adopt. In addition, many of these techniques are quite rapid and require very little material, both important attributes at the early stage of target identification and selection.

PROTEIN SECONDARY STRUCTURES AND MOTIFS

The secondary structure of a protein is determined by the set of dihedral angles (ϕ, φ), which define the spatial orientation of the peptide backbone, and the presence of specific hydrogen bonds. When the backbone dihedral angles have repeating values, the peptide forms regular secondary structures. The principal geometry for the α-helix is $\phi \sim -60°$ and $\varphi \sim -45°$ with hydrogen bonds from the NH of the fifth residue in the chain to the C=O group on the first residue, or between residues i and ($i + 4$). It forms a compact, rod-like structure with 3.6 amino acids (1.5 Å) per repeating unit and a radius, excluding side chains, of 2.3 Å. The α-helix forms a right-handed screw as one looks down its axis from the amino terminal end. In soluble proteins, the average length of a helix is 11 residues, corresponding to three turns. Because all of the backbone amide groups are involved in intrachain hydrogen bonds, the interactions of helices with other peptide domains or small molecules occur exclusively through side-chain interactions. The α-helix is a favored struc-

ture in nonpolar solvents due to the extensive hydrogen bonded network, which removes the peptide backbone from hydrogen bonding to the solvent. Amphipathic helices, in which one face of the helix is polar and the other nonpolar, are a common protein motif, in which nonpolar residues interact between adjacent helices or at a membrane interface. The dihedral angles of the β-sheet are $\phi \sim -130°$ and $\varphi \sim 120°$, forming an extended structure with some right-handed twist. Hydrogen bonds between protein chains, oriented either in a parallel or in an antiparallel fashion, stabilize β-sheets in proteins and protein complexes.

Turns allow a protein to fold back on itself and are stabilized by a hydrogen bond that holds the ends together. They are classified according to the number of residues involved in the hydrogen-bonded structure. The peptide backbone in a β-turn forms a rough plane that contains the intramolecular hydrogen bond. However, the amide bond between the ($i + 1$) and the ($i + 2$) residues is perpendicular to this plane and must hydrogen-bond elsewhere. In addition, the side chains of turns project outward. Thus, β-turns are often found on the surface of proteins where hydrogen bonding with the solvent is favorable. Because of their surface location, turns play major roles in molecular recognition. Unordered or random structure is generally defined as a conformation that is not helix, sheet, or turn.

CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism is observed when molecules absorb left and right circularly polarized light to different extents. The amide chromophore of the peptide bond in proteins dominates their CD spectra below 250 nm. In an α-helical protein, a negative band near 222 nm is observed due to the strong hydrogen-bonding environment of this conformation. This transition is relatively independent of the length of the helix. A second transition at 190 nm is split into a negative band near 208 nm and a positive band near 192 nm. Both bands are reduced in intensity in short helices. The CD spectra of β-sheets display a negative band near 216 nm, a positive band between 195 and 200 nm, and a negative band near 175 nm. However, the position and magnitude of these bands is variable, resulting in less accurate predictions for β-structure than for α-helices by CD. A review of various methods to estimate the conformation of proteins from CD data (1) has recently been published. CD is particularly useful for measuring the temperature dependence of protein secondary structure (Fig. 1).

Sample Preparation

Sample preparation for CD spectroscopy is similar to that employed for UV-Vis spectrophotometry in the same wavelength range. Cuvettes are made of fused
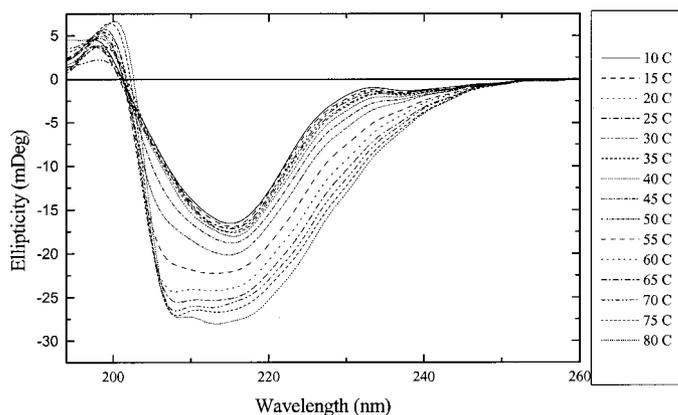
FIG. 1. CD spectrum of interleukin-1 as a function of temperature (R. A. A. Atkinson and J. T. Pelton, unpublished data).

quartz and are either circular or rectangular. Circular cuvettes are generally used for room temperature measurements. Temperature may be controlled either with a recirculating water bath and appropriate cuvettes or through a Peltier device that requires rectangular cuvettes. Path lengths are generally 0.1 to 1 mm to minimize absorption from solvent. Hellma (Forest Hills, NY) offers cells that require only the front face of the cuvette to be filled. Only 0.35 mL of sample is required in the 1 mm path length cuvette.

The selection of buffer and solvent solutions is critical. It is best to check the spectrum of the solvent alone before using it. However, one may be given a sample without complete specification of the solvent. The best practice is to exchange the sample buffer with a suitable buffer prior to recording the spectra. Any buffer or excipient that absorbs in the UV (such as imidazole) or is optically active (such as nucleotides) should be removed from the sample. Chloride salts must be strenuously avoided as they absorb in the far UV. Chaps and octylglucoside at low concentrations are the preferred detergents. Triton detergents tend to oxidize rapidly and form UV-absorbing materials.

Special care must be taken for the optimal concentration of material and to reduce turbidity of samples. The maximum absorbance of the sample in the wavelength range of interest should not exceed 2. The optimal absorbance (at which the signal to noise ratio is a maximum) is 0.87, but lower concentrations can be used. An excellent spectrum can be obtained on as little as 0.01 mg of protein. Generally, CD measurements are made on relatively dilute protein solutions (0.01–0.2 mg/mL). The advantage of dilute solutions is that concentration-dependent aggregation of the sample is minimized. Care must be taken in handling such dilute solutions. Contact with glass pipettes and test tubes may cause protein loss due to adsorption to the glass surfaces.

Accurate protein concentrations are essential for determining the distribution of secondary structures in a sample. The extinction coefficient of protein at 280 nm may be calculated from its content of tryptophan (ϵ = 5690 M^{-1}cm^{-1}), tyrosine (ϵ = 1280 M^{-1}cm^{-1}), and cystine (ϕ = 120 M^{-1}cm^{-1}) in 6 M GdnHCl (2). This calculation is generally accurate to about 5%, even when the measurements are made in nondenaturing buffers (3). Given the strong dependence of the calculated secondary structure values on protein concentrations, it is useful to bracket the protein concentration by calculating at 95, 100, and 105% of the measured values.

Spectral Acquisition

In a commercial CD instrument, monochromatic, linearly polarized light passes through a piezoelastic modulator (PEM) to give circularly polarized light. Preferential absorption of the right or left circularly polarized component by the sample reduces the intensity of the corresponding transmitted light. The output of the photomultiplier tube consists of a DC component (E_A) corresponding to the average amount of transmitted light and an AC component (E_s) related to the modulation of the light. The DC component is kept constant by varying the voltage applied to the photomultiplier (PM) tube. The AC component is proportional to the CD signal. The sample compartment and optics are purged with nitrogen in commercial instruments. A detector sensitive to the blue edge of UV spectrum is selected to give a usable range of 170 to 700 nm. In modern instruments, the lower limit is generally constrained by the sample and solvent, rather than the optics.

The CD spectrum is recorded in millidegrees of ellipticity as a function of wavelength. For proteins, the quantity of interest is the mean residue ellipticity (in deg.cm^2/dmol): $[\theta] = \theta/10 \cdot cl,$ where c is the molar concentration of amino acid residues and l is the pathlength of the cell in centimeters. Mean residue ellipticity values may be converted to units of molar circular dichroic absorption ($\Delta\epsilon$) by dividing by 3298.

A two-point calibration is required as the relative magnitudes of the long-wavelength bands to the short-wavelength bands strongly influence the outcome of a secondary structural analysis (4). Ammonium camphor sulfonate is used as the calibration standard [$\Delta\epsilon_{290.5}$ = 2.36 and $\Delta\epsilon_{192.5}$ = −4.90, for the acid (5)]. Thus, a 0.06% ammonium d-10-camphor sulfonate solution (stored for up to 8 months in a glass container in the refrigerator) gives 18.7 mdeg at 290.5 nm and −38.9 mdeg at 192.5 nm in a 1-mm cuvette.

The signal to noise ratio (S/N) of the measurement is improved by employing multiple, slow scans with long time constants (response times). The improvement in S/N is equivalent whether the number of scans (N) or the response time (R) is increased, as $S/N \propto (R \times N)^{1/2}$.

To minimize distortion of the measured spectrum, the response-wavelength width (the product of the scanning speed in nm/s and response time in s) should be less than one-tenth of the half-height width (generally about 15 nm for a protein spectrum). High-quality spectra can be obtained with a response time of 4 s and a scan rate of 20 nm/min. Collection of 4 spectra from 260 to 178 nm takes less than 20 min. Postrun noise reduction by data-smoothing algorithms or Fourier methods should not be necessary if the sample is sufficiently concentrated and the operator is patient. Rapid scans collected during a kinetic process are generally smoothed for presentation purposes.

Spectral Analysis

Simple inspection of CD spectra reveals information about the structural class of the protein (6). Proteins with substantial α-helical structure (all-α proteins) exhibit pronounced negative bands at 222 and 208 nm and a positive band between 190 and 195 nm. The 208-nm band is larger than that at 222 nm in $\alpha + \beta$ proteins and smaller in α/β proteins. The all-β proteins lack the characteristic 222/208 peaks. An estimate of the α-helical content may be made from the mean residue ellipticity at 222 nm, fraction α-helix $= -([\theta] + 3000)/33,000$ (7), or at 208 nm, α-helix $= -([\theta] + 4000)/29,000$ (8).

A more complete analysis of the secondary structure assumes that the spectrum is a linear sum of the spectra of the individual secondary structures, plus contributions from aromatic chromophores and noise. The earliest methods used multiple linear regression to fit the experimental CD spectrum to a small set of reference spectra. The reference spectra are determined from the CD spectra of model polypeptides or globular proteins with known crystal structures (see (1) for review). In constrained least-square fits, the sum of all structures is constrained to 100%. In unconstrained fits, the coefficients are normalized to 100% after fitting. An assumption in all of these methods is that the solution and crystallographic secondary structures of the proteins are identical.

An alternative method is to extract basis spectra by singular value decomposition (SVD), a type of multicomponent analysis. A complete set of information is given by a set of eigenvectors that can reconstruct the original CD spectra within experimental error. Most of the information is found in the region 195–250 nm, consistent with the observation that truncated data sets are generally accurate for α-helical structures. A simple matrix multiplication application of this method has been published (9). Good data must be collected to at least 184 nm for a successful fit. A similar method (convex constraint analysis, CCA) that deconvolutes a set of spectra with known secondary structures into components fares better with truncated

data sets (10), but is unable to resolve conformational assignments in an unknown protein structure.

The most accurate modern methods employ selection criteria in which only spectra similar to that of the experimental spectrum form the basis set. The Provencher and Glöckner (11) method analyzes a CD spectrum as a sum of 16 reference spectra of proteins of known secondary structure. Reference spectra that have shapes similar to that of the experimental spectrum are given greater weight in the fitting procedure (regularization). This allows the total number of reference spectra to be increased over that of earlier least-squares procedures and dramatically improves the fit. The fit is constrained to physically realistic values of secondary structures.

The problem of fitting spectra is threefold: (1) what proteins should be used as the basis data set; (2) how are their contents of secondary structure determined; and (3) how can component spectra be extracted out of the CD spectra. Greenfield (1) recently applied a wide range of methods of analysis to a consistent set of 16 proteins. Secondary structures were assigned to the test proteins by means of the X-ray coordinates of their crystal structures using the method of Kabsch and Sander (12). All of the analysis methods provide accurate values for α-helical structure. Estimates of β-structures are more variable, but the best methods (SELCON and CONTIN) can predict β-sheets with a high degree of accuracy. Both SELCON (13) and CONTIN (11) can be used with limited data sets (truncated at 200 nm) without a dramatic loss of accuracy. This means that problematic samples that absorb in the far UV may be analyzed.

FOURIER TRANSFORM INFRARED (FT-IR) AND RAMAN SPECTROSCOPY

Vibrational spectra may be obtained either by infrared absorption (IR) or Raman scattering spectroscopy. The two techniques provide complementary information. IR results from the absorption of energy by vibrating chemical bonds (primarily stretching and bending motions). Raman scattering results from the same types of transitions, but the selection rules are somewhat different so that weak bands in the IR may be strong in the Raman and vice versa. Raman spectra are reported as the difference between the incident and scattered radiation frequencies. These frequencies depend on the types of bonds and their modes of vibration. Characteristic groups of atoms give rise to vibrational bands near the same frequency regardless of the molecule in which they are found. The precise wavenumbers of bands within this range depend on inter- and intramolecular effects, including peptide-bond angles and hydrogen-bonding patterns. Thus, vibrational spectra can be used to estimate the secondary struc-

Endothelin-1 in H₂O
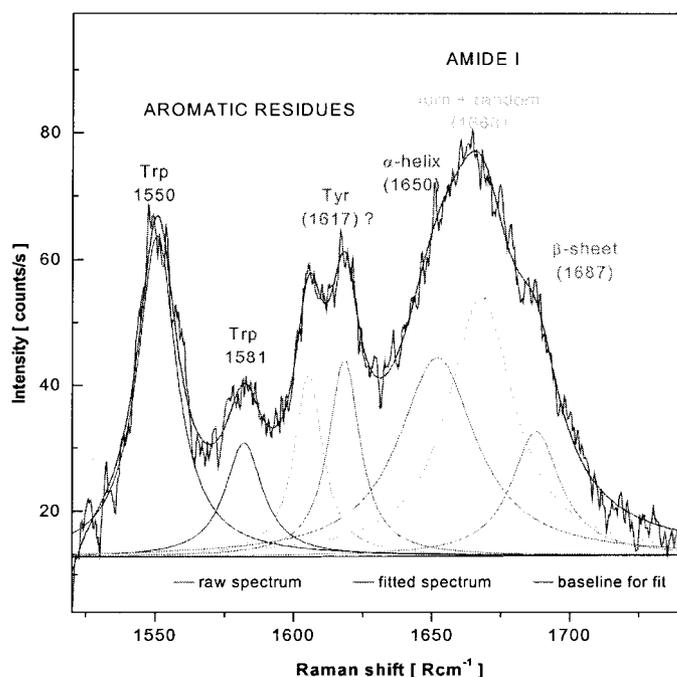
FIG. 2. Laser Raman spectrum of crystaline endothelin-1 (ET-1). Fourier self-deconvolution of the amide I band was followed by curve fitting to generate the underlying components. Integration of band areas provides an estimate of the secondary structure (C. Brockel and J. T. Pelton, unpublished data).

ture of proteins by inspection of the frequencies at which the amide bonds absorb infrared radiation. The application of infrared and Raman spectroscopies to protein secondary structure has undergone a renaissance with the development of Fourier transform spectrometers and improved computers.

Nine normal modes are allowed for the amide band of proteins. These are called A, B, and I-VII in order of decreasing frequency. The amide bands I (80% C=O stretch, near 1650 cm^{-1}), II (60% N–H bend and 40% C–N stretch, near 1550 cm^{-1}), and III (40% C–N stretch, 30% N–H bend, near 1300 cm^{-1}) are generally employed to study protein structure. The amide I and III bands have appreciable Raman intensities with visible light excitation. In practice, the amide I band in FT-IR and the amide I and III bands in Raman are primarily used to assign secondary structures to proteins. Identification of particular frequencies with secondary structures has been made by reference to spectra of homo-polypeptides and proteins with primarily α-helical or β-sheet structures, theoretical calculations (normal mode analysis), synthetic peptides, and proteins with known three-dimensional structures. Raman also provides information on aromatic residues in the region below about 1620 cm^{-1} (Fig. 2).

Sample Preparation

Infrared spectra with sufficiently high signal to noise ratios for measurements of proteins in solutions are problematic on all but Fourier transform instruments. However, such instruments are routinely used in modern analytical chemistry and biophysics labs and are the least expensive instruments of those that can provide information on protein secondary structures. Less care must be taken for salt solutions than in CD and turbid samples are not a problem. The limitation is that very short pathlengths must be used (typically 6 μm) with high protein concentrations (typically 10–20 mg/mL) in order to prevent the water band from entirely swamping the detection system. As in CD, the intensity of the spectrum depends on the number of chromophores, which are the peptide bonds.

Demountable cells with spacers are routinely used for FT-IR measurements because of the ease of cleaning. Care must be taken in assembling the cell to prevent leaks that change the volume of the cell and cause problems in subtracting the buffer spectrum. To minimize leaks, spacers with no signs of wrinkles or pits should be used. Cells may be made from a variety of insoluble materials. ZnSe and CdTe have high refractive indices which reduce light throughput. CaF$_2$ and BaF$_2$ are preferred even though they tend to cloud when in contact with aqueous solutions. This clouding can be removed by polishing the windows. Attenuated total reflectance (ATR) has also been applied to IR measurements of hydrated films of proteins. The band positions and characteristics of the spectra are similar, but the Beer's Law assumption used in transmission measurements may not apply (14). Raman spectra can be collected directly in aqueous solutions using simple glass capillaries.

A primary difficulty in IR measurements is the absorption of water at 1650 cm^{-1}, in the middle of the amide I band. Deuterated water (D$_2$O) may be used instead of water, but the spectrum may be more complicated if isotope exchange is not complete, as deuteration displaces the frequencies of the absorption bands (Table 1). Thus, careful subtraction of the buffer spec-

TABLE 1

Principal Amide I Frequencies Characteristic of Protein Secondary Structures

Conformation	H₂O	D₂O
α-helix	1650–1657	1647–1654
Antiparallel β-sheet	1612–1640; 1670–1690 (weak)	1628–1635
Parallel β-sheet	1626–1640	
Turn	1655–1675 1680–1696	
Unordered	1640–1651	1643

trum is necessary before meaningful data can be obtained. In contrast to IR, water gives rise to a weak Raman spectrum. However, it is more sensitive to fluorescence artifacts and sensitive samples may decompose in the intense laser excitation beam. Infrared absorbing contaminants and counterions may also produce artifacts in the spectra. Of particular concern with synthetic peptides is trifluoroacetate, which is used in purification. This salt gives a strong band near 1673 cm^{-1}, which may be erroneously assigned to the amide I mode. Clearly, there is no substitute to knowing the source and history of each sample prior to making a measurement if incorrect interpretations of the data are to be avoided.

Spectral Acquisition

FT-IR instruments employ an interferometric pattern of IR radiation that passes through the sample to a detector. The interferometric transmitted light in the time domain is converted to the frequency domain by applying the fast Fourier transform (FFT). Division of this single beam spectrum by that of the empty compartment gives the final spectrum. Averaging a large number of scans (often over 1000) may reduce instrumental noise, so instrumental stability is important. S/N should be greater than 500 prior to spectral manipulations. Raman spectra are obtained by irradiating a sample with monochromatic radiation. In early studies, UV lasers were used. Currently, the 514.5-nm line of an argon-ion laser is generally employed. The magnitude of the frequency shift is independent of the wavelength of the irradiating source. The spectrum of the scattered radiation is typically measured at a 90° angle from the incident light. Raman lines are very weak, constituting 0.001% or less of the intensity of the source. Fourier transform Raman spectrometers dramatically improve signal/noise ratio and are sensitive enough to use longer wavelength excitation sources, such as a Nd:YAG laser, that minimize fluorescent artifacts.

Subtraction of the buffer in IR spectroscopy requires close attention to pathlength, temperature, and sample compartment atmosphere. Inefficient purging of the sample compartment with dry air or nitrogen may result in residual water vapor that is not properly subtracted. One solution to this problem is to use a sample shuttle that alternately places the sample and buffer cells in the beam. After acquisition of the spectrum of the empty cell, buffer, and protein, a spectrum of the cell after flushing with buffer should be obtained to check for the presence of any adsorbed protein. Because it is not possible to precisely match the sample and buffer collection conditions, the buffer spectrum is multiplied by a scaling factor in an iterative process until the baseline in a region of the spectrum where

peaks are not expected ($1730–2100 \text{ cm}^{-1}$) is flat (15). While Raman spectra also require subtraction of solvent and buffer components prior to analysis, the generally weak nature of these contributions makes the task relatively straightforward.

Spectral Analysis

Resolution enhancement. The primary problem in IR spectral analysis of proteins is that the bands are a complex composite of overlapping component bands that represent different structural elements. The width of the component bands is usually greater than the separation between the maxima of adjacent peaks and cannot be resolved by simple inspection of the spectra. While resolution enhancement of IR and Raman spectra does not increase the instrumental resolution, it does lead to separation of the underlying band structure and permits individual components to be visualized. The successful application of resolution enhancement techniques requires careful attention to the elimination of water vapor and random noise from the spectrum. The latter can produce artifacts indistinguishable from amide bands after application of enhancement techniques. The absence of artifacts must be verified by a careful inspection of the resolution-enhanced protein spectrum adjacent to the amide I band ($1700–1800 \text{ cm}^{-1}$) where peaks are not generally expected.

Fourier self-deconvolution (FSD) provides band narrowing through multiplication of the Fourier transform by a line-shape function and an apodization function. The result yields a Fourier transform of narrower bandwidth. Generally, a value of 13 cm^{-1} for the full bandwidth at half height (FWHH) and a resolution enhancement factor (K) of 2.4 are adequate (16). Deconvolution using a too narrow FWHH will not separate all the widest bands, while a value that is too large will overemphasize the spectrum's narrower components leading to artifacts and possible misinterpretation. For this reason, second derivative spectra are commonly employed to validate the FSD parameters.

Amide I bands. Table 1 summarizes the IR frequencies generally diagnostic of specific protein secondary structures. Proteins known to adopt an α-helical conformation have strong amide I bands between 1650 and 1655 cm^{-1}. The hydrogen-bonding strengths in β-sheets are more variable due to their flexibility and tendency to twist. A strong band between 1612 and 1640 cm^{-1} and a weaker band about 1685 cm^{-1} are commonly observed for β-sheets, although weak bands at somewhat lower frequencies ($1665–1670 \text{ cm}^{-1}$) have also been observed. Identification of these bands is aided by H→D exchange experiments as the amide backbone protons in β-sheets exchange quite slowly at room temperature and the difference in frequency be-

tween the amide I and amide I′ (in D_2O) bands is large (Table 1). Only a single strong band at 1670 cm^{-1} is observed in Raman spectra. The amide I band of turns overlaps those of helices and sheets making assignment of this component solely from amide I frequencies difficult. Unordered or random structure is generally assigned to the band near 1665 cm^{-1} in the Raman and 1645 cm^{-1} in the IR. The latter is close to the frequencies associated with α-helix.

Amide II bands. Due to the large contribution of the N–H bend to the amide II mode, deuteration results in a substantial shift to lower frequency ($\sim$1460 cm^{-1}). In IR, a strong amide II band is observed at 1540–1550 cm^{-1} and a weaker shoulder at 1510–1525 cm^{-1}. Peptides and proteins with an antiparallel β-sheet structure have strong amide II bands between 1510 and 1530 cm^{-1}; a parallel β-sheet structure is found at somewhat higher frequencies (1530–1550 cm^{-1}). Inspection of both the amide I and II bands generally provides little help in distinguishing between turn and sheet conformation. In favorable situations, however, the amide II band may assist in assigning random structure and allow a more accurate estimate of the helix and random components.

Amide III bands. Like the amide II band, deuterium substitution of the hydrogen atom shifts the amide III band to lower wavenumber (960–1000 cm^{-1}). In the infrared, the amide III band is normally quite weak and occurs in a region of mixed vibrations (CH bending, tyrosine and phenylalanine ring vibrations) that are not readily correlated to protein secondary structure. In the Raman spectrum the amide III band is stronger. Although it appears in a region of the spectrum that contains a number of unrelated vibrational bands, the amide III components can be identified by their shift in D_2O. When used in combination with the amide I band, their assignment may permit a distinction between β and disordered structure that is generally not possible based on the amide I band alone.

Quantitation of secondary structures. Once the component amide bands have been identified on the basis of the resolution-enhanced and second derivative spectra, quantification assumes that the extinction coefficients for the different structural elements are the same. Thus, intensities are proportional to the fraction of each secondary structure. Since the bandshapes after resolution enhancement are distorted, enhanced spectra cannot be used directly to estimate secondary structure. Rather, an iterative process of adjusting the positions, intensities, and shapes of the component bands to the experimental spectrum is employed. If the final positions of major bands change considerably from initial values, the spectra are carefully reinspected and resolution enhancement is repeated with somewhat different parameters.

A major difficulty in using curve-fitting procedures to estimate secondary structure by FT-IR is the inability to cleanly resolve disordered and helical bands in H_2O. In addition, the band at 1660–1663 normally assigned to turns is often found to reflect the presence of some helix or irregular loop structure. This has led to an overestimation of the turn content in several proteins. Curve fitting of the amide I IR band gives reasonably accurate estimates for β-sheet content. However, while the Raman amide I band can usually differentiate between helical and β or disordered structure, it is poor at distinguishing between the latter two.

Multivariate statistical methods have also been used for analysis of protein secondary structure by employing spectra from proteins of known structure. These methods are similar to those used for CD spectroscopy and include classical least squares (CLS), in which pure-component spectra are extracted from the data; partial least squares (PLS), which builds a small basis set composed of linear combinations of the original calibration spectra (15); and factor analysis (17). These methods employ the full spectrum and can be applied simultaneously to amide I and amide II (1500–1600 cm^{-1}) bands. Inclusion of the amide III (1200–1300 cm^{-1}) bands tends to increase the error of prediction, probably due to the mixed nature of bands in this region.

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

NMR spectroscopy is one of only two techniques that can provide detailed structural information about macromolecules at atomic resolution. This detailed view of molecular structure results from a laborious examination of a number of conformationally sensitive parameters and the application of distance geometry programs to provide high-resolution structures of peptides and proteins to about 30,000 MW. However, unlike the organic chemist, who has long characterized small molecules by applying empirical "rules" associating the chemical shift with structure and conformation, the use of chemical shift as a tool to understand biological conformation has not been widely employed. The major difficulty has been a poor understanding of the link between chemical shifts and structural parameters. While the theoretical difficulties remain largely unsolved, there now exists a large body of NMR chemical-shift data for peptides and small proteins that can be used to develop empirical relationships.

In an early statistical study, Szilagyi and Jardetzky identified a significant correlation between αH chemical shifts and helical and β-sheet structures (18). In the absence of other effects, helical conformations produce up-field shifts while β-structures shift the α proton downfield. A smoothed plot of $^1H\alpha$ chemical shifts as a function of sequence can be used to readily identify

regions of secondary structure (19). A closely related method assigns a chemical shift index (CSI) to each residue in a protein, by comparison with a table of chemical shifts corresponding to random structure (20). Regions where the CSI is clustered with negative values are assigned as α-helical; those with positive values are assigned to β-structure.

In contrast to optical methods for determination of protein structure, NMR provides information on the location of secondary structural elements within the protein sequence. Oldfield has even suggested that sufficiently accurate data may be used to predict the three-dimensional structure of the protein from chemical-shift data alone (21). In order to apply chemical-shift information to predictions of secondary structure, the chemical shifts must be assigned to particular residues in the protein. This is a tedious task that requires the measurement and analysis of 2-D and often 3-D spectra. In addition, the limit of about 30K for a protein that produces sufficiently narrow lines seriously hampers the general application of the method. However, the chemical-shift index serves as a useful check on further model refinement in high-resolution NMR studies of small proteins.

SELECTED APPLICATIONS

Applications of secondary structural measurements cover a broad expanse of systems and questions of relevance to biochemistry. In this section, we provide a brief review of selected applications that have recently appeared in major journals. It is not our intention to be either inclusive or exhaustive, but to simply illustrate some of the applications that may be of interest to readers and to provide an entry point into the literature.

The interaction of ligands with proteins will generally stabilize a conformation somewhat different from that of the unbound protein structure. Such changes have been observed in CD spectra with binding of ATP and Mg^{2+} to Rad51 (22) and in IR spectra following binding of ATP or ADP to GroEL (23). In both cases, the effects of the nucleotides are small and not accompanied by dramatic changes in secondary structure. In contrast, interfacial binding of proteins can have dramatic effects on secondary structure. The α-helicity of α-synuclein is increased from 3 to 80% upon binding to phospholipid bilayers (24). Fleury and co-workers (25) combined Raman spectroscopic data with statistical prediction techniques to construct a model of secondary structure distribution in topoisomerase I. Binding of a sequence-specific oligonucleotide was accompanied by a conformational transition evident in both CD and Raman spectra. Addition of Zn^{2+} to secretin displaces specific ^1H chemical shifts in the N-terminal region, consistent with metal binding to His_1 (26). The helical

portion of the hormone was identified by its NMR chemical-shift index and corresponds closely with the amount of α-helix measured by CD.

Thermal denaturation measurements by CD are commonly used to assess the stability of proteins and protein complexes and are especially useful in assessing the effects of amino acid mutations on protein structure. However, thermal denaturation is irreversible and cannot be used to measure the thermodynamic parameters of protein folding and unfolding. For this, denaturants are commonly used. Fersht and co-workers (27) used guanidinium hydrochloride (GdnHCl) to initiate unfolding of the tetramerization domain of the tumor-suppressor protein p53 (p53tet). Loss of secondary structure was monitored by measurements of the molar ellipticity of the proteins at 222 nm after rapid mixing by stopped flow. Similarly, Zhang and Matthews (28, 29) studied the urea-induced folding and unfolding kinetics of p21ras. Removal of GDP and Mg^{2+} reduced the stability of the protein, but had only a minimal effect on its secondary structure. Large changes in Raman bands characteristic of protein denaturation are observed between 30 and 40°C in both the α- and β-subunits of the *Oxytricha nova* telomere-binding protein (30). The spectra and denaturation profile of the combined subunits do not differ from those of the sum of the constituents, indicating that there is no specific interaction between subunits.

The location of binding sites on proteins has been investigated by site-directed mutagenesis, synthetic peptides, and specifically labeled residues. The mechanism of acid-induced release of ligand from the macrophage scavenger receptor was elicited by replacing acidic amino acids with neutral amino acids (31). Analysis of the structures of the mutants demonstrated that Glu_{242} was sufficient to induce a pH-dependent conformational change of the α-helical coiled coil domain. Johnson (32) showed that peptides corresponding to the putative membrane-binding domain of CTP:phosphocholine cytidylyltransferase bound to vesicles as α-helices with selectivity for anionic lipids. Specific isotope labeling of Tyr residues in rhodopsin with deuterium and measurements of FT-IR difference spectra suggested that Tyr_{185} and Pro_{186} contribute to a hinge that facilitates α-helix movement during photoactivation of the protein (33).

Protein-protein interactions may be observed by measurements of the stability of protein complexes or by the effects of association on secondary structure. The stability of protein complexes of SNARE proteins with vesicle-associated membrane protein (VAMP) and synaptosome-associated protein-25 (SNAP-25) assessed by thermal denaturation using CD is relatively independent of the various combinations of SNARES (34). This suggests that the specificity of membrane fusion is not encoded by the interactions between

SNAREs. The binding of a synthetic peptide from the major calmodulin-binding domain of cyclic nucleotide phosphodiesterase with calmodulin is accompanied by formation of an α-helix in the presence, but not in the absence, of calcium (35).

Finally, optical methods have been used to probe the structure of membrane receptors. The resolution-enhanced amide I band of the nicotinic acetylcholine receptor after exposure to D_2O is consistent with an α-helical conformation of the protein that slowly exchanges protein protons with solvent deuterium (36). These data strongly support an α-helical conformation for the transmembrane domain of the receptor. Consistent with these data, peptides corresponding to the transmembrane segments of the receptor adopt α-helical conformations when reconstituted into lipid vesicles (37). Receptor studies are not limited to purely structural investigations. By IR spectroscopy a peptide corresponding to the $K_v3.4$ channel ball domain adopts a partial β-sheet structure in the presence of anionic lipids (38). These data were used to refine a molecular model that may explain the role of this domain in inactivation of the receptor.

PERSPECTIVES AND SUMMARY

The three-dimensional structures of a relatively large number of soluble protein molecules have been solved, revealing a wide range of structures and a smaller number of structural motifs. A far greater number of amino acid sequences have been determined directly with isolated proteins or predicted from the sequences of the encoding cDNA. The fundamental dogma of protein structure is that the amino acid sequence defines the structure of the protein. It is generally believed that proteins fold to attain the lowest energy structure that may be attained for a particular sequence. Thus, it has been a major goal of computational chemists to predict the structure of a protein from its sequence. This program has been carried out with varying degrees of success, accelerating the pace of discovery through molecular biological techniques, which are considerably more rapid than full-scale determinations of protein crystal structures. However, one must be cautious in applying computational techniques and not become overly zealous in defining protein structures in the absence of experimental confirmation. Fold recognition is the first step in defining a model for a target protein. Knowing what parts of a protein adopt particular secondary structures is critical for almost all protein-modeling approaches.

Often the availability of proteins in high purity or as crystals suitable for diffraction experiments is a limiting factor for detailed biophysical studies. Several classes of structural interest consist of complex mixtures of macromolecules, which are not amenable to

such high-resolution approaches. In both cases, spectral and theoretical techniques can provide much needed information about the modes and types of binding interactions of potential drugs with their receptors, leading to the design of new compounds potentially useful as therapeutic agents.

REFERENCES

1. Greenfield, N. J. (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **235,** 1–10.

2. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6,** 1948–1954.

3. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326.

4. Toumadje, A., and Johnson, W. C., Jr. (1993) Effects of relative band intensity on prediction of protein secondary structure from CD. *Anal. Biochem.* **211,** 258–260.

5. Chen, G. C., and Yang, J. T. (1977) Two-point calibration of circular dichrometer with d-10-camphorsulfonic acid. *Anal. Lett.* **10,** 1195–1207.

6. Manavalan, P., and Johnson, W. C., Jr. (1983) Hydrophobic character of amino acid residues in globular proteins. *Nature* **305,** 831–832.

7. Morrisett, J. D., David, J. S., Pownall, H. J., and Gotto, A. M., Jr. (1973) Interaction of an apolipoprotein (apoLP-alanine) with phosphatidylcholine. *Biochemistry* **12,** 1290–1299.

8. Greenfield, N., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8,** 4108–4116.

9. Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **155,** 155–167.

10. Perczel, A., Park, K., and Fasman, G. D. (1992) Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: A practical guide. *Anal. Biochem.* **203,** 83–93.

11. Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20,** 33–37.

12. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637.

13. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209,** 32–44.

14. de Jongh, H. H., Goormaghtigh, E., and Ruysschaert, J. M. (1996) The different molar absorptivities of the secondary structure types in the amide I region: An attenuated total reflection infrared study on globular proteins. *Anal. Biochem.* **242,** 95–103.

15. Dousseau, F., and Pezolet, M. (1990) Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods. *Biochemistry* **29,** 8771–8779.

16. Byler, D. M., and Susi, H. (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25,** 469–487.

17. Lee, D. C., Haris, P. I., Chapman, D., and Mitchell, R. C. (1990) Determination of protein secondary structure using factor analysis of infrared spectra. *Biochemistry* **29,** 9185–9193.

18. Szilagyi, L., and Jardetzky, O. (1989) Alpha-proton chemical shifts and secondary structure in proteins. *J. Magn. Reson.* **83,** 441.

19. Pastore, A., and Saudek, V. (1990) The relationship between chemical shift and secondary structure in proteins. *J. Magn. Reson.* **90,** 165–176.

20. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4,** 171–180.

21. Oldfield, E. (1995) Chemical shifts and three-dimensional protein structures. *J. Biomol. NMR* **5,** 217–225.

22. Namsaraev, E. A., and Berg, P. (1998) Interaction of Rad51 with ATP and Mg^{2+} induces a conformational change in Rad51. *Biochemistry* **37,** 11932–11939.

23. von Germar, F., Galan, A., Llorca, O., Carrascosa, J. L., Valpuesta, J. M., Mantele, and Muga, A. (1999) Conformational changes generated in GroEL during ATP hydrolysis as seen by time-resolved infrared spectroscopy. *J. Biol. Chem.* **274,** 5508–5513.

24. Davidson, W. S., Jonas, A., Clayton, D. F., and George, J. M. (1998) Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273,** 9443–9449.

25. Fleury, F., Ianoul, A., Kryukov, E., Sukhanova, A., Kudelina, I., Wynne, J., Bronstein, I. B., Maizieres, M., Berjot, M., Dodson, G. G., Wilkinson, A. J., Holden, J. A., Feofanov, A. V., Alix, A. J., Jardillier, J. C., and Nabiev, I. (1998) Raman and CD spectroscopy of recombinant 68-kDa DNA human topoisomerase I and its complex with suicide DNA-substrate. *Biochemistry* **37,** 14630–14642.

26. Carpenter, K. A., and Schiller, P. W. (1998) Aggregation behaviour and $Zn2+$ binding properties of secretin. *Biochemistry* **37,** 16967–16974.

27. Mateu, M. G., Sanchez, D. P. M., and Fersht, A. R. (1999) Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. *Nature Struct. Biol.* **6,** 191–198.

28. Zhang, J., and Matthews, C. R. (1998) Ligand binding is the principal determinant of stability for the p21(H)-ras protein. *Biochemistry* **37,** 14881–14890.

29. Zhang, J., and Matthews, C. R. (1998) The role of ligand binding in the kinetic folding mechanism of human p21(H-ras) protein. *Biochemistry* **37,** 14891–14899.

30. Laporte, L., Stultz, J., and Thomas, G. J., Jr. (1997) Solution conformation and interactions of alpha and beta subunits of the Oxytricha nova telomere binding protein: Investigation by Raman spectroscopy. *Biochemistry* **36,** 8053–8059.

31. Suzuki, K., Yamada, T., and Tanaka, T. (1999) Role of the buried glutamate in the alpha-helical coiled coil domain of the macrophage scavenger receptor. *Biochemistry* **38,** 1751–1756.

32. Johnson, J. E., Rao, N. M., Hui, S. W., and Cornell, R. B. (1998) Conformation and lipid binding properties of four peptides derived from the membrane-binding domain of CTP:phosphocholine cytidylyltransferase. *Biochemistry* **37,** 9509–9519.

33. DeLange, F., Klaassen, C. H., Wallace-Williams, S. E., Bovee-Geurts, P. H., Liu, X. M., DeGrip, W. J., and Rothschild, K. J. (1998) Tyrosine structural changes detected during the photoactivation of rhodopsin. *J. Biol. Chem.* **273,** 23735–23739.

34. Yang, B., Gonzalez, L. J., Prekeris, R., Steegmaier, M., Advani, R. J., and Scheller, R. H. (1999) SNARE interactions are not selective. Implications for membrane fusion specificity. *J. Biol. Chem.* **274,** 5649–5653.

35. Yuan, T., Walsh, M. P., Sutherland, C., Fabian, H., and Vogel, H. J. (1999) Calcium-dependent and -independent interactions of the calmodulin-binding domain of cyclic nucleotide phosphodiesterase with calmodulin. *Biochemistry* **38,** 1446–1455.

36. Methot, N., and Baenziger, J. E. (1998) Secondary structure of the exchange-resistant core from the nicotinic acetylcholine receptor probed directly by infrared spectroscopy and hydrogen/deuterium exchange. *Biochemistry* **37,** 14815–14822.

37. Corbin, J., Methot, N., Wang, H. H., Baenziger, J. E., and Blanton, M. P. (1998) Secondary structure analysis of individual transmembrane segments of the nicotinic acetylcholine receptor by circular dichroism and Fourier transform infrared spectroscopy. *J. Biol. Chem.* **273,** 771–777.

38. Abbott, G. W., Mercer, E. A., Miller, R. T., Ramesh, B., and Srai, S. K. (1998) Conformational changes in a mammalian voltage-dependent potassium channel inactivation peptide. *Biochemistry* **37,** 1640–1645.