# The Analysis of Raman Optical Activity Spectra of Proteins

by

## Christopher D. Syme

A thesis presented in partial fulfilment for
the degree of Doctor of Philosophy in the
Faculty of Science of the University of Glasgow

# Acknowledgements

# Table of Contents

# Abstract

Measurement of the Raman optical activity (ROA) spectra of biomolecules has become an experimental possibility due to significant advances in the available technology, and its successful implementation into the ROA instruments at the University of Glasgow. The ease with which the ROA spectra of biological molecules can be successfully measured lends itself perfectly to the ever-growing demand for biomolecular structural information, especially in the context of proteomics and the Human Genome Project. ROA spectroscopy is able to probe the chiral peptide backbone of proteins, and as such the ROA spectrum of a protein contains a wealth of structural information from within the whole molecule, across the whole vibrational spectrum. As well as containing detailed information from specific structural elements such as sections of secondary structure and motifs, the ability of ROA to see the molecule as a whole also enables the global fold of the protein to be deduced from the ROA spectrum.

The development of the analysis of ROA spectra has largely been based upon the correlation of ROA spectra of proteins of known structure with structural information from alternative sources, chiefly X-ray crystallography and multidimensional nuclear magnetic resonance (NMR). As the database of ROA spectra of polypeptides and proteins has grown, it has been possible to tighten up the assignment of ROA spectral bands and band patterns to aspects of known structural content. With a basis for the correlation between the ROA spectrum and the known crystal structure (or NMR structure) being well established, it is possible to interpret the ROA spectra of proteins that do not have (for whatever reason) well defined structures. This means that ROA spectroscopy can provide invaluable structural information for proteins that are precluded from analysis by other techniques, and also cast new light on the structures of proteins that have not been well defined.

In order to fully interpret an ROA spectrum of a protein, it is necessary to be familiar with protein structure and the ROA experiment as a whole. Analysing an ROA spectrum is a detailed and highly subjective process. Depending on the experience of the analyst, the information contained within the spectra can be

extracted readily or not so readily. For this reason, it would be desirable to develop a technique that is capable of interpreting not only individual spectra, but also whole data sets in a manner that is independent of the analyst, and therefore independent of any preconceptions (or inexperience) the analyst may have.

This project presents an up-to-date collection of newly obtained ROA spectra of a large number of proteins across a range of structural class types. In addition, the statistical technique of principal component analysis (PCA) has been used as a tool for the analysis of this new data. It is hoped that the result of this work will provide a basis for the future analysis of protein ROA spectra that is both mathematically rigorous and convenient.

# Chapter One
# Overview

## 1.1 Introduction

Raman optical activity (ROA) measures vibrational optical activity as a small difference in the Raman scattering of a chiral molecule in right and left circularly polarized incident laser light. The first genuine observations of vibrational optical activity in a chiral molecule in the liquid phase were made in 1973 (Barron et al., 1973) using the ROA technique. The technique has now been developed to the point where it is possible to measure on the same instrument high-quality ROA spectra of almost any chiral molecular system from the smallest such as CHFClBr, to the largest such as intact viruses (Barron & Hecht, 2000).

Raman spectroscopy is a form of vibrational spectroscopy complementary to infrared spectroscopy. Whereas infrared (IR) spectroscopy involves the absorption of infrared radiation by the unexcited vibrational states within molecules, Raman spectroscopy involves the inelastic scattering of visible photons to yield the vibrational spectrum of a molecule. The Stokes Raman scattering event involves the interaction of a visible photon of energy $\hbar\omega$ (where $\omega$ is the angular frequency) with a molecule, leaving the molecule in an excited vibrational state of energy $\hbar\omega_v$. The photon loses energy to the molecule such that there is a shift to lower angular frequency $\omega - \omega_v$ of the scattered photon. By using a visible spectrometer to analyze the scattered light, the complete vibrational spectrum of the molecule can be obtained. Raman spectroscopy has been applied to many biochemical problems since it is straightforward to obtain the complete vibrational spectrum (~100 to 4000 cm$^{-1}$) in one measurement and water, the natural biological medium, is an excellent solvent for Raman studies.

This being the case, ROA studies may also benefit from these useful experimental features. But it is the phenomenon of vibrational optical activity that adds the extra dimension of stereochemical sensitivity, and hence the usefulness of the technique for studying molecules that contain chiral centres, such as proteins.

Raman scattering is a weak phenomenon (~1 photon per million), and ROA is a further three orders of magnitude weaker than that. For this reason, it has taken many years to develop an instrument and experimental protocol that is capable of studying aqueous solutions of biological molecules, since large fluorescent backgrounds and water bands previously swamped the ROA signal from the chiral peptide backbone of the protein molecules.

The usefulness of ROA in protein science becomes truly apparent when compared to the key experimental techniques it complements, namely X-ray crystallography and multidimensional nuclear magnetic resonance (NMR). X-ray crystallography and multidimensional NMR are powerful, atomic-resolution techniques that are capable of providing the most in-depth probe of protein structure currently possible, both of which are capable of pinpointing the positions of each atom that the protein molecule is composed of. However, both of these high-resolution techniques suffer from major drawbacks that have until now (and also for the foreseeable future) placed severe limits on the range of proteins that they are able to study. Diffraction methods such as X-ray crystallography (and neutron scattering) require an ordered crystalline sample. Many proteins simply do not crystallise, and many others that do are unable to form crystals that are suitable for analysis by diffraction methods. NMR requires suitably soluble proteins, and analysis is limited by the size of the protein. At present, 2D NMR is limited to around 14 kDa, whereas 3D and 4D NMR (which involve isotopic substitution) can raise this limit to at best 27 kDa. These latter techniques are also prohibitively expensive, and have a theoretical limit of 30 to 50 kDa.

The lower resolution techniques that are currently widely used to study protein structure include vibrational spectroscopies such as Fourier transform infrared (FTIR), Raman and vibrational circular dichroism (VCD); electronic spectroscopy such as ultraviolet absorbance and fluorescence and ultraviolet circular dichroism (UVCD), mass spectrometry and a host of other techniques such electron paramagnetic resonance (EPR) and dynamic light scattering. It is important to point out that no one technique mentioned here would be used as a sole probe of protein structure. For the purpose of this discussion, it is the alternative vibrational spectroscopic techniques that are of most interest for comparison to ROA. Conventional Raman spectroscopy

of proteins suffers from the fact that the largest contribution to the Raman spectrum comes from the bulky side chain groups from the individual amino acids, which has the effect of swamping conformational information from the peptide backbone of the molecule and hence the three dimensional structure of the protein. Another widely used spectroscopic technique is UVCD. This involves the differential absorption of circularly polarized radiation. The peptide backbone $\pi^* \leftarrow n$ and $\pi^* \leftarrow \pi$ transitions in the far UV range from ~190-220 nm yield information pertaining to α-helix, β-sheet and random coil (Woody, 2000; Rogers et al., 1997) Also, transitions in the near UV range from ~250-300 nm give information from aromatic side chains, useful in the study of the denaturation of proteins. However, a major disadvantage of UVCD is the sparseness of detail contained in the spectra. Broad and frequently overlapping bands sometimes make attempts at quantitative analysis unreliable. However, recent work using principal component analysis on a number of protein UVCD spectra has shown the potential to separate out the individual factors that influence the shape of protein UVCD spectra (Pancoska et al., 1991) and this provided the initial inspiration for similar work on protein ROA spectra discussed as part of this thesis.

The low resolution techniques benefit from the fact that sample preparation and the experimental techniques involved are relatively routine. ROA is an excellent example of this fact. In order to study the solution structure of a protein by ROA, it is only necessary to prepare a solution of suitably high concentration. In general this is a straightforward process that is easy to replicate to a high level of accuracy. Similarly, sample preparation for UVCD experiments is also straightforward. Proteins are simply dissolved in an aqueous buffer with no limit to the size of the protein molecules. Compared to X-ray crystallography, these techniques have the potential to provide 'high throughput' sources of structural information, a useful requirement in the field of proteomics.

The experimental technique that is now in practice at Glasgow University has yielded a data set that represents the most detailed and reproducible protein ROA spectra to date. As a result of the consistently superior quality of the data, it has been possible to construct a meaningful basis for the analysis of the spectra, such that the spectrum of an unknown protein may be interpreted with a high level of confidence.

A major aim of this project was to solidify ROA band assignments and investigate the relationship between the tertiary fold of proteins and the appearance of band patterns within ROA spectra. To this end, it was necessary to obtain the ROA spectra of a set of proteins that survey as many structural classes and fold types possible.

## 1.2 Theory

**Introduction**

Chiral molecules exhibit optical activity such that they are capable of rotating the plane of incident polarized light. In the ROA experiment, optical activity of the target molecule results in a non-zero difference in Raman scattering between right and left circularly polarized light. The basic ROA observable is $I^R - I^L$, where $I^R$ and $I^L$ are the Raman scattered intensities in right and left circularly polarized incident light (Figure 1.1).

R     L

incident beam with
angular frequency $\omega$

$$I^R - I^L \neq 0$$

Raman component of scattered beam
with angular frequency $\omega - \omega_v$
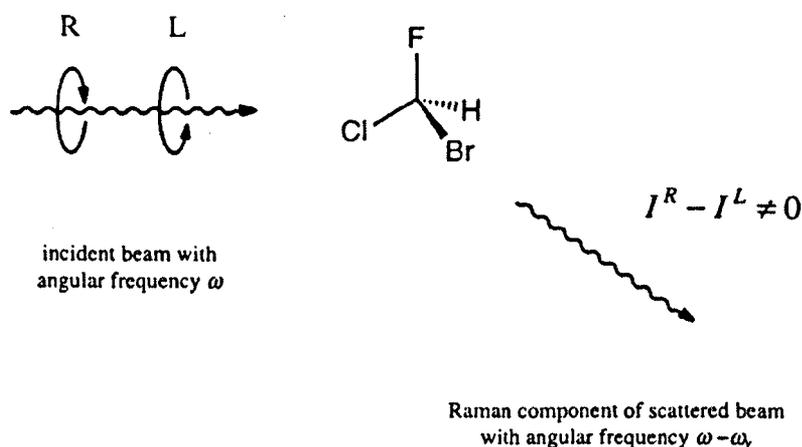
**Figure 1.1** The basic ROA experiment measures a small difference in the intensity of Raman scattering from a chiral molecule in right and left circularly polarized light

**Basic theory of ROA**

The fundamental scattering mechanism responsible for ROA was discovered by Atkins and Barron (1969) who showed that interference between waves scattered via the polarizability and optical activity tensors of the molecule yields a dependence of the scattered intensity on the degree of circular polarization of the incident light and to a circular component in the scattered light (Barron et al., 2000). Subsequently,

a more definitive version of the theory was formulated (Barron & Buckingham, 1971) and introduced the definition of the dimensionless circular intensity difference (CID),

$$\Delta = (I^R - I^L)/(I^R + I^L) \tag{1.1}$$

as an experimental quantity. In terms of the electric dipole–electric dipole molecular polarizability tensor $\alpha_{\alpha\beta}$ and the electric dipole–magnetic dipole and electric dipole–electric quadrupole optical activity tensors $G'_{\alpha\beta}$ and $A_{\alpha\beta\gamma}$ (Buckingham, 1967; Barron, 1982; Polavarapu, 1998), the CIDs for forward (0°) and backward (180°) scattering from an isotropic sample for incident transparent wavelengths much greater than the molecular dimensions are as follows (Barron, 1982; Hecht & Barron, 1990; Nafie & Che, 1994; Polavarapu, 1994):

$$\Delta(0^\circ) = \frac{4\left[45\alpha G' + \beta(G')^2 - \beta(A)^2\right]}{c\left[45\alpha^2 + 7\beta(\alpha)^2\right]} \tag{1.2a}$$

$$\Delta(180^\circ) = \frac{24\left[\beta(G')^2 + 1/3\beta(A)^2\right]}{c\left[45\alpha^2 + 7\beta(\alpha)^2\right]} \tag{1.2b}$$

where the isotropic tensor invariants are defined as

$$\alpha = \frac{1}{3}\alpha_{\alpha\alpha}, \quad G' = \frac{1}{3}G'_{\alpha\alpha} \tag{1.3}$$

and the anisotropic tensor invariants are defined as

$$\beta(\alpha)^2 = \frac{1}{2}(3\alpha_{\alpha\beta}\alpha_{\alpha\beta} - \alpha_{\alpha\alpha}\alpha_{\beta\beta}) \tag{1.4a}$$

$$\beta(G')^2 = \frac{1}{2}(3\alpha_{\alpha\beta}G'_{\alpha\beta} - \alpha_{\alpha\alpha}G'_{\beta\beta}) \tag{1.4b}$$

$$\beta(A)^2 = \frac{1}{2}\omega\alpha_{\alpha\beta}\varepsilon_{\alpha\gamma\delta}A_{\gamma\delta\beta} \tag{1.4c}$$

Using a simple bond polarizability theory of ROA for the case of a molecule composed entirely of idealized axially symmetric bonds, the relationships $\beta(G')^2 = \beta(\alpha)^2$ and $\alpha G' = 0$ are found (Barron & Buckingham, 1974; Barron, 1982). Within this model, isotropic scattering makes zero contribution to the ROA intensity which is generated exclusively by anisotropic scattering, and the CID expressions reduce to

$$\Delta(0°) = 0, \tag{1.5a}$$

$$\Delta(180°) = \frac{32\beta(G')^2}{c\left[45\alpha^2 + 7\beta(\alpha)^2\right]} \tag{1.5b}$$

Therefore, unlike the conventional Raman intensity that is the same in the forward and backward directions, the ROA intensity is maximized in backscattering and zero in forward scattering. It follows that the best experimental strategy for the measurement of ROA intensity is a backscattering arrangement, as is the case with the current instrument (Hecht et al., 1989; Hecht et al., 1999).

A simple realization of the bond polarizability model in which chiral structure composed of two achiral anisotropic groups held in a chiral twisted arrangement
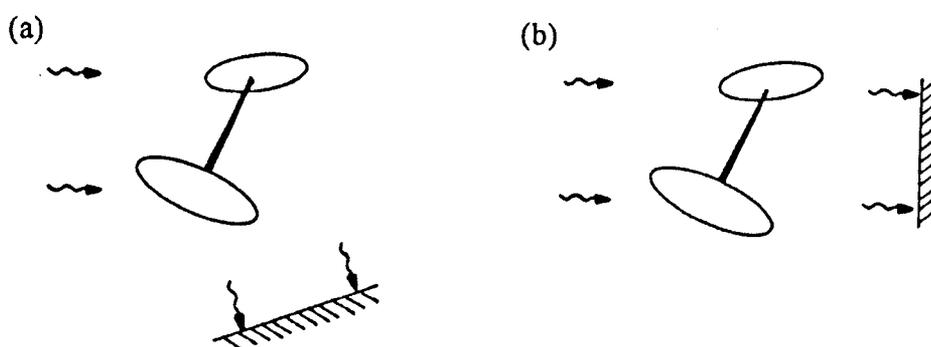


**Figure 1.2.** (a) Generation of ROA from interference between two scattered waves from two achiral groups held in a twisted chiral conformation. (b) No ROA is generated in the forward direction as there is no phase difference between the two waves.

provides a simple picture of the generation of ROA (Figure 1.2). This so-called two-group model illustrates that in light scattered away from the forward direction, there is interference between two photons independently scattered from the two groups, which is sufficient to provide chiral information.

**ROA as a probe of chiral biomolecules**

Biomolecules such as polypeptides and proteins have become the major focus of attention for study by ROA due to the presence of chiral centres throughout their structures, which leads to many structure-sensitive ROA bands. A major advantage of ROA is that the largest signals are associated with vibrational coordinates which sample the most rigid and chiral parts of the structure i.e. the polypeptide backbone. As such, the resulting vibrational spectra largely reflect secondary structure and fold information as opposed to just information from the amino acid side chains. ROA spectra do contain a few distinct bands from amino acid side chains which can sometimes be useful, but in general these bands do not obscure peptide backbone bands as in conventional Raman spectra of biomolecules.

## 1.3  Experimental

**The instrument**

The current instrument at Glasgow, developed by Barron and Hecht is capable of performing measurements of ROA spectra of complex biological molecules, a goal that has been realized through the continuous development of optical and electronic components of the instrument over many years (Hecht et al., 1999). The instrument is a high performance and finely tuned Raman spectrometer, capable of obtaining a sufficiently high signal to noise ratio of the extremely weak ROA signal, such that a useful protein ROA spectrum can be measured within a matter of hours, depending upon the nature of the sample. Figure 1.4 illustrates the optical layout of the instrument.  The backscattering arrangement is essential for the measurement of protein ROA studies, as discussed in the previous section. This arrangement is facilitated by using a mirror with a hole, placed at 45° to the path of the cylinder of
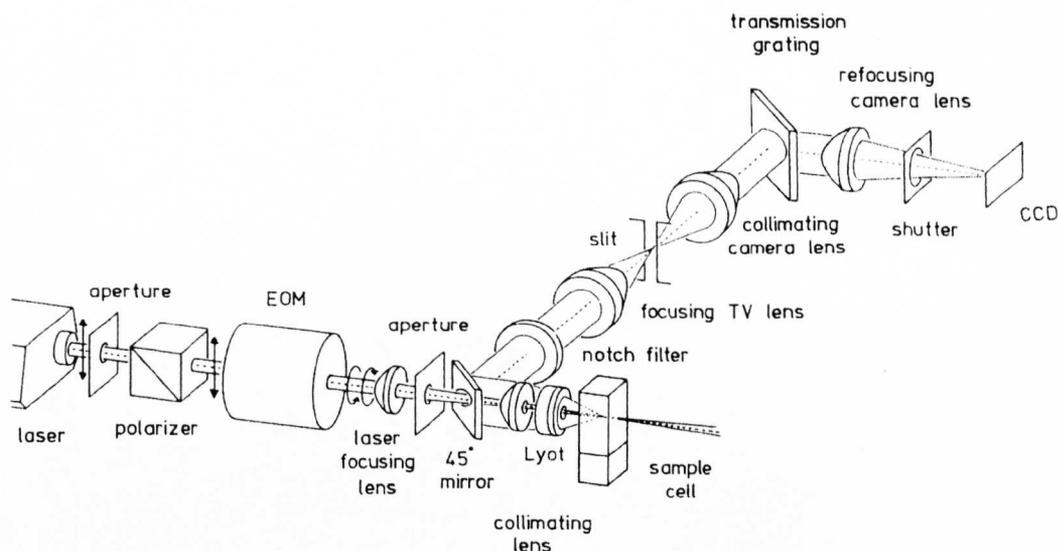
**Figure 1.4** Optical layout of the current ROA instrument at Glasgow University

collimated backscattered light. This enables the light scattered by the sample at 180° (backscattering) to be diverted into the detector part of the instrument.

## Laser light and circular polarization

The excitation source is a continuous-wave argon ion laser (Spectra-Physics Model Stabilite 2017-05S) that operates at an excitation wavelength of 514.5 nm (visible green light). The reason that this wavelength is used can be partially explained by considering the $1/\lambda^4$ Rayleigh scattering law, such that higher wavelength (red light or IR radiation) would provide insufficient scattering for ROA to be measured efficiently. Using lower wavelength radiation (blue light) would boost the signal, but would also greatly enhance the fluorescence associated with complex biological molecules meaning that the weak ROA signal would be swamped (Hecht & Barron, 1994).

The linear polarization of the laser light is cleaned up by an air-spaced calcite Glan-Taylor prism polarizer (LEYSOP, Model GT12) (Hecht et al., 1992) before it passes into a customized electro-optic modulator (EOM) (LEYSOP). The EOM contains a Z-cut potassium di-deuterium phosphate crystal that acts as a quarter-wave plate retarder which converts linearly polarized light into circularly polarized light.

8

When a pre-determined quarter-wave voltage is applied to the gold ring electrodes attached to front and back faces of the crystal, a longitudinal electric field is generated such that the crystal retards a component of linearly polarized light by a quarter-wave, generating circularly polarized light as illustrated in Figure 1.5. By alternating the direction of the voltage applied and hence the direction of the electric field across the crystal, the EOM modulates the state of the light between right and left circular polarization. An essential element in the routine collection of ROA spectra is the precise control of the EOM. The voltage being applied across the crystal is generated by a high-voltage linear differential amplifier (LEYSOP, Model LDA 50) (Hecht et
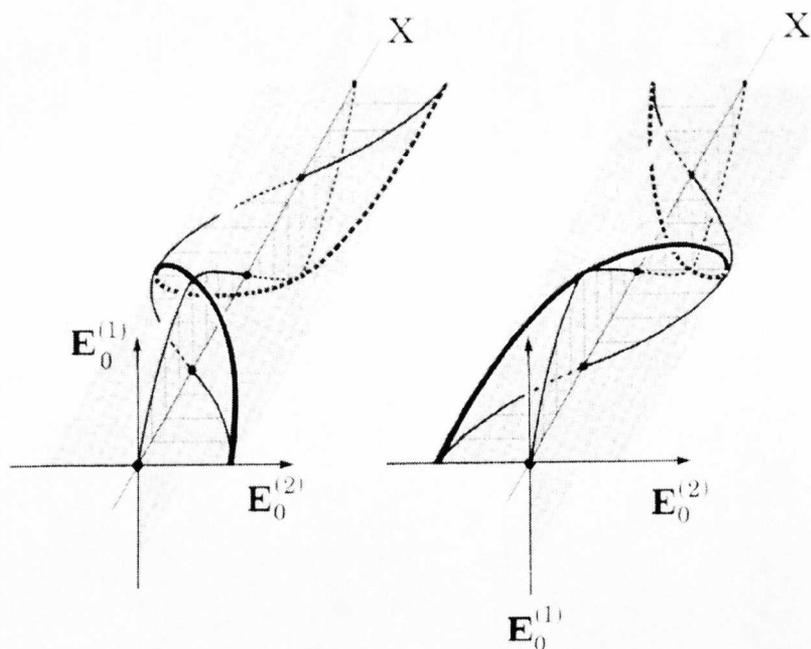


**Figure 1.5** Graphical representation of right and left circular polarization states generated by two orthogonal plane waves with phases offset by + and − 90° respectively

al., 1992) which provides a highly stable bipolar square wave. Precise control of the voltage being applied to the crystal is made possible by a hand-adjustable voltage control, which allows the voltage to be minutely offset, positively or negatively, from the pre-determined values. This enables the correction of baseline drifts that occur due to birefringence effects from different samples, and changes in the sample during an experiment. The frequency of modulation between polarization states, and hence the acquisition time of each pair of Raman spectra, is controlled by a computer connected to the amplifier.

## The rotating half-wave plate

Immediately after the EOM in the beamline, a half-wave plate mounted in a rotatable stage has been added. This important instrumental development was implemented at the start of the project in the Autumn of 1998 and has had a marked effect on the ease of data collection and has reduced problems associated with linear contaminants within the laser light. The spectra presented in this thesis represent the first data collected using a rotating half-wave plate, and comparison with previously acquired spectra reveals a distinct improvement in not only the signal to noise ratio, but also in the overall reliability of the experiment, making a significant contribution to the goal of routine protein ROA data collection. Although a number of factors have contributed to the general improvement in the acquired data, the addition of the rotating half-wave plate has had the greatest single effect. By rotating the half-wave plate, the small linear contaminants in the circular polarization states produced by the EOM, and which are different in the right and left states, are continuously rotated through 360° and hence the artefacts they produce average to zero.

## The optics block

Once converted to circularly polarized light by the EOM, the laser light is focused into the sample cell by a 150 mm focal length plano-convex laser focusing lens. A small aperture of ~1 mm is used to block any side beams from the laser line that may cause artefacts in the ROA spectrum. Before reaching the sample cell, the light passes through an optical block that contains the 45° mirror, a collimating lens and the Lyot depolarizer. Each component in the block has a ~2 mm hole drilled in the middle so that the light can pass unhindered into the sample cell. During the set up of the instrument, it is essential that these components are aligned exactly, so that the edges of the holes do not clip the incident laser beam.

## The sample cell

The sample is contained within a micro-fluorescence fused quartz silica cell made to our design by Optiglass, with internal dimensions 4 x 4 x 25 mm, capable of containing volumes up to 150 µl. Previous ROA experiments have been carried out using larger cells with thicker glass walls and larger internal volumes, factors which add to difficulties with artefact generation and a requirement for greater amounts of

sample per experiment respectively. These newer, more compact cells also come anti-reflection coating on the external faces that decrease stray light problems. Good alignment of the sample cell is the single most important experimental factor in the day to day operation of the experiment. The sample cell is held in a mount that has adjustable screws that allow the alignment of the cell to be precisely controlled in the x-, y- and z- directions, such that minute adjustments can be made to the position and orientation of the sample relative to the beam. Only once the acquisition of an ROA spectrum has begun does it become clear whether the cell alignment is correct or not. If the cell alignment is incorrect, it is possible that light that has been reflected back from the cell walls can enter the collection optics and produce artefacts and baseline slopes. The focus of the laser beam must also fall within the body of the sample liquid such that the cone of backscattered light does not encounter the cell walls on exit as this too can produce problems with artefacts (Hecht & Barron, 1995).

**Collection of backscattered light**

The cone of backscattered light from the sample contains light that has been scattered via both the Rayleigh and Raman scattering mechanisms. The vast majority of this is Rayleigh (or elastic) scattering, which has the same frequency as that of the incident light. This light is removed by a holographic Super Notch-Plus filter (Kaiser Optical Systems) before it is allowed to enter the collection optics. The notch filter is an important component as stray Rayleigh light adversely affects the signal-to-noise ratio and introduces baseline slopes into the ROA spectra. The remainder of the scattered light is from Raman scattering. The backscattered Raman light passes through the components of the optical block in reverse order to that describes before (now coming from the opposite direction). The light passes through the Lyot depolarizer, which scrambles any residual linearly polarized light. The Lyot is made from two calcite plates, one twice as thick as the other, held together with optical cement with their axes at 45° to each other. It is mounted in a rotatory stage that permits rotation such that the optimal position can be found on the basis of experimental trial and error. The rotatory stage has markings that allow the Lyot to be moved by any desired amount, and the optimum angles of performance noted.

The backscattered Raman light is collimated and reflected into the collection optics by the 45° mirror at this point. At the entry to the collection optics is the holographic notch filter, which has a transmission efficiency of greater than 80% at wavelengths of interest in a typical protein ROA spectrum (~600-1800 $cm^{-1}$). The efficiency of the notch filter decreases as the angle of incidence of the beam moves away from 90°. However, it was discovered that by tilting the notch filter slightly away from exactly perpendicular to the on-coming cylinder of depolarized Raman light, a hitherto undiscovered source of stray light was diverted away from entering the collection optics, resulting in a further improvement in artefact reduction. The fact that some of the bright cylinder of light incident upon the shiny surface of the notch filter was reflecting directly back into the optical block had not been noticed before. Once the Rayleigh light has been removed, the collimated backscattered light is focused through an 8mm entrance slit of a single-stage stigmatic spectrograph (Kaiser Optical Systems) by a 75mm f/1.4 TV camera lens (Nikon). The f-number (f/1.4) is the ratio of the focal length to the diameter of the lens (f/D = 1.4). A low f-number (such as for this lens) indicates the ability of the lens to allow more light to pass and therefore reach the detector. An electronically controlled shutter is placed in front of the slit, which opens at the start of a spectral run and closes at the end.

**The spectrograph**

Within the spectrograph, an 85mm f/1.4 TV camera lens (Nikon) collimates the Raman light on to a holographic transmission grating. This novel diffraction grating is based on volume-phase technology (Tedesco et al., 1993) and folds the optical path through 90° with > 80% efficiency for unpolarized light. This is a marked improvement compared with previous designs based on ion-etched gratings with ~ 60% efficiency. The spatial frequency of the volume holographic diffraction grating is equivalent to 2400 grooves $mm^{-1}$ resulting in a reciprocal dispersion of 3.0 nm $mm^{-1}$. The dispersed beam is refocused by another 85 mm f/1.4 TV camera lens, identical to that placed directly before the holographic diffraction grating, and passes through an exit port.

**The detector**

The last component of the instrument is the detector. The CCD (charge-coupled device) camera (Wright Instruments) is fixed to the exit port of the spectrograph by mounting flanges that allow for the precise positioning of the camera. The camera has a back-thinned spectroscopic-grade CCD chip of 1024 columns and 256 rows of individually spaced square pixels. Back-thinning increases the quantum efficiency of the CCD when compared with conventional non-thinned devices since the photo-induced charges have less distance to travel before being detected (Benyon et al., 1980). The chip is made from light sensitive silicon oxides which produce electronic charge when impinged upon by Raman scattered light. The whole chip is cooled to ~200 K in order to minimize the dark current (or thermally induced signal), resulting in an improved signal to noise ratio. To maximise spectral coverage, the chip is aligned with its long axis parallel to the direction of dispersion of the spectrograph. This allows for a spectral coverage of ~2500 cm$^{-1}$ for Stokes Raman shifts for 514.5 nm laser excitation. The two-dimensional CCD image is converted into a linear array by a serial pixel binning method (Tedesco et al., 1993). The binning process involves the addition of the signal from adjacent pixels on the CCD chip, or after readout. On-chip binning is preferred as it reduces the total readout noise associated with the CCD output amplifier (Sweedler et al., 1988). Overall, the CCD detection system is highly effective with a quantum efficiency in the region of ~80% in the 450 to 650 nm range.

**Computer control**

The separate acquisition of Raman spectra in left and right circularly polarized light requires synchronization of the voltage applied to the EOM and the collection of data from the CCD. This is achieved by using a computer equipped with a customized version of the LabCalc software package (Galactic Industries). The software is programmed to accept the first acquisition to correspond to the right circularly polarized state. The spectrum is displayed on the monitor which enables the acquisition time required to maximize the parent Raman signal to be determined prior to ROA acquisition. This time defines how long each acquisition (in left and right CPL) is to take, and also therefore, how often the signal is to be read from the chip. Once the ROA acquisition has begun, the first spectrum of each pair is collected and

stored in the memory of the PC. When the second of each pair is collected, the software then calculates the difference between the two (i.e $I^R - I^L$). This signal is then displayed as the ROA spectrum. This process is repeated until the ROA signal has built up until a sufficiently good signal to noise ratio is achieved, which typically takes between 6 and 24 hours.
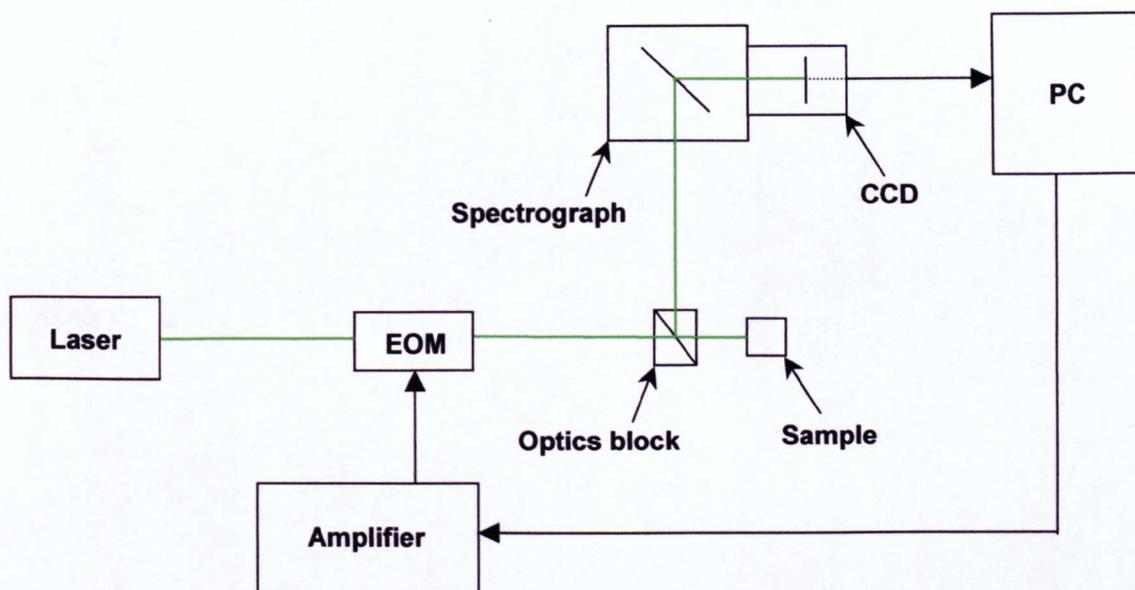


**Figure 1.6** Schematic representation of the instrument with integrated computer control

## 1.4  Experimental procedure

**Sample preparation**

To facilitate ROA measurements of proteins, samples are normally prepared at relatively high concentration, typically between ~50-100 mg/mL in clean, dry miniature test tubes. A small amount of potassium iodide (~8 mM) is sometimes added to the buffer solutions used in preparing samples which quenches fluorescence from impurities. Activated charcoal is added to remove fluorescent impurities and is later removed by centrifugation. The remaining solution is added to a Hamilton syringe fitted with a 0.22 µm Millipore filter, and transferred directly into quartz microfluorescence cells. The cells may be centrifuged for a short period immediately prior to mounting the cell in the instrument to minimize interference from small particulates (such as dust and small protein aggregates) that may remain. The exterior

walls (both front and back) of the sample cells are cleaned by using lens cleaning paper with a little methanol. Once the cell is in the instrument, the remaining fluorescent background of the sample is "burnt down" by allowing the samples to equilibrate in the laser for several hours prior to the acquisition of ROA data.

## Experimental procedure

With the sample prepared and the cell placed in the instrument, the fluorescent background may be monitored by looking at the parent Raman spectrum. In order to attain the greatest efficiency from the experiment, it is necessary to attempt to maximise the time per acquisition (acquisition time), whilst ensuring that the relative intensity of the parent Raman spectrum stays below ~25,000 ADC counts. Above this level, the CCD chip becomes saturated and the response non-linear. By maximizing the acquisition time, the amount of acquisition pairs is minimized and therefore the noise associated with the readout procedure is kept to a minimum.

Once the ROA spectral acquisition process is underway, it is possible to control the drift towards a negative or positive bias caused by slight imperfections in the circularly polarized state of the light by using the rotating half wave plate as discussed earlier. In addition, voltage offsets may be applied from the amplifier to counteract any visible drifting. The duration of the experiment largely depends on the purity of the sample and the accuracy of the experimental set-up. Typical ROA acquisition times are between ~5-24 hours, with straightforward protein samples yielding high quality (high signal to noise ratios) data within ~5-8 hours.

# Chapter Two
# Raman Optical Activity of Proteins

## 2.1 Introduction

Proteins are macromolecules that carry out a vast array of vital biological functions. They are essentially long chains of amino acids which, once coded for by the DNA sequence of the host and assembled in the cell, fold in on themselves to adopt complex three-dimensional structures. The function of a protein is determined by its structure, and therefore detailed knowledge of protein structure is essential in understanding their role in biological function and dysfunction. Proteins are made from the 20 naturally occurring amino acids listed in Appendix 3. As well as an abbreviated name, each amino acid is given a single letter code that enables long chains of amino acids to be written in a more convenient form. Amino acids link together by means of peptide linkages as shown in Fig 2.1. Short chains of amino acids are known as peptides, and it is common to refer to the chain of peptide links in proteins as the peptide backbone of the molecule.

Raman optical activity offers a new perspective on protein structure. The major advantage of ROA over alternative spectroscopic techniques is that it is sensitive to chirality. As will be discussed in more detail later in this chapter, the peptide backbone that determines the overall shape (and hence function) of a protein contains chiral centres throughout, making proteins ideal targets for analysis by ROA.

## 2.2 Protein Structure

**Introduction**

Protein structure was first described in a hierarchical fashion by Linderstrøm-Lang in 1952. The first level of protein structure is simply the sequence of amino acids, known as the primary structure. Fig, 2.1 illustrates how amino acids are joined together, creating chiral alpha-carbons with bond angles relative to the plane of the bulk of the amino acid molecule denoted by $\phi$ and $\psi$. The range of $\phi$ and $\psi$ angles

commonly observed in proteins is restricted to discrete values. This is a result of the fact that individual amino acids have different physico-chemical properties and hence each have difference propensities towards favouring certain conformations in the folded protein molecule.
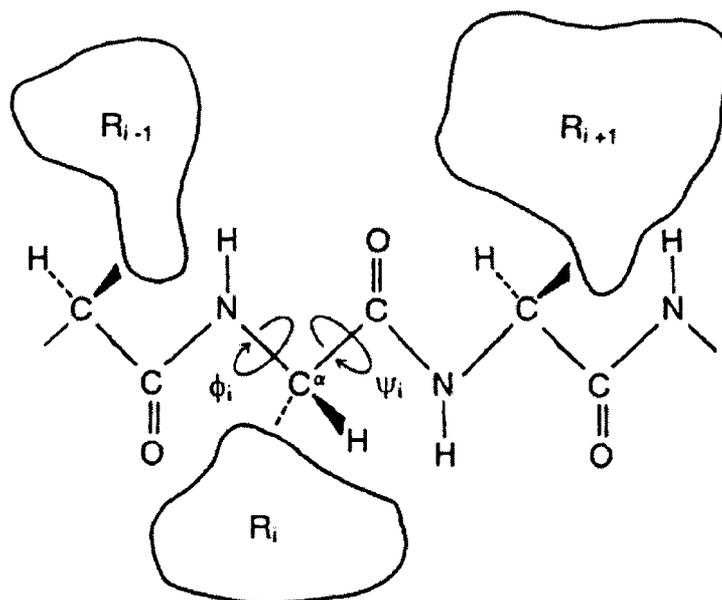


**Figure 2.1.** Diagrammatical representation of peptide linkage between amino acids. The chiral alpha carbon is $C^\alpha$ and bond angles $\phi$ and $\psi$ are illustrated. Side chains are denotes $R_i$ etc.

## Secondary structure

Sequences of amino acids with similar propensities order themselves into regular structures. These structures are known as secondary structural elements and have specific topological and physical properties

The $\alpha$-helix is the most common form of secondary structure (Pauling & Corey, 1951) with approximately 31% of all residues in proteins participating in $\alpha$-helices. The $\alpha$-helix has 3.6 residues per turn and is stabilised by hydrogen bonding between the backbone carbonyl oxygen of one residue and the backbone NH of the fourth residue along the helix. Amino acids in ideal $\alpha$-helices have $\phi$ and $\psi$ angles of approximately -60° and -50°. Although both right- and left-handed $\alpha$-helices are found in proteins, the vast majority are right-handed as they have more favourable steric interactions between amino acid side chains. The dipole moments across the peptide

bond line up in an α-helix to form a helical macrodipole. Interactions involving these helical macrodipoles are important factors in stabilising protein structures.

α-helices are frequently amphipathic. This means that the distribution of hydrophilic and hydrophobic amino acids along the helix results in a hydrophobic face and a hydrophilic face. This is a very important characteristic for protein structure. Most proteins exist in an aqueous environment (with the exception of membrane proteins) and it is a general rule of protein structure that proteins have hydrophobic cores. In proteins, α-helices interact in such a way that the hydrophobic areas form hydrophobic interactions with each other, and the hydrophilic areas are exposed to the solvent. Certain amino acids have a distinct preference to form α-helices; for example alanine, glutamic acid, leucine and methionine are good helix-forming residues and proline, glycine, tyrosine and serine are helix-breaking residues.

The second most common element of secondary structure in proteins is the β-sheet (Pauling et al., 1951). β-sheet is formed from several individual β-strands that are distant from each other along the primary protein sequence. β-strands are usually 5-10 residues long and are in a fully extended conformation possessing $\phi$ and $\psi$ angles that occupy a wide range of values in the fully allowed region of the Ramachandran plot (see Table 2.1) The individual strands are aligned next to each other in such a way that the peptide bond carbonyl oxygens hydrogen bond with neighbouring NH groups. β-sheet stabilises as a hydrogen bonding network is built up.

Two types of connection topology are seen in β-sheet, the most stable being anti-parallel β-sheet. Most of the β-proteins studied by ROA have this type of β-sheet. In anti-parallel sheets, the β-strands are connected sequentially. Parallel β-sheet is less stable due to the hydrogen bonds not being optimally aligned. Parallel sheets are formed from segments of peptide backbone distantly connected by other types of secondary structure. In β-sheet, the amino acid side-chains of successive residues project alternately from either side of the β-sheet. It is common to find one side of the sheet predominantly hydrophobic and one side hydrophilic. Such hydrophilic faces frequently make contact via hydrophobic bonds. Almost all β-sheet found in proteins

display a pronounced right hand twist. Parallel β-sheet is observed in the α/β class of proteins, and a handful of proteins have a mixture of both anti-parallel and parallel β-sheet.

The aforementioned elements of secondary structure are connected to each other by loop segments of polypeptide. Generally, α-helices and β-sheets form the stable hydrophobic core of the protein. The connecting loops are usually to be found on the surface of the structure. As they are solvent exposed, they are rich in polar and charged amino acids which hydrogen bond to solvent water molecules as opposed to participating in concerted hydrogen-bonding patterns. Related proteins from different organisms frequently display a very similar hydrophobic core region but have large variation in the surface loops. The ability to detect structural differences that arise due to the subtle yet important differences in loop structure is one of the major advantages of ROA spectroscopy over conventional spectroscopic techniques.

Finally, random coil is the term used for segments of polypeptide chain that do not form regular secondary structures. Random coil is somewhat of a misnomer as it results from a balance of interactions between amino acid side chains and solvent and interactions between side-chains. The predominant hydrogen bonding pattern in random coil is between the polypeptide chain and water, with intrachain hydrogen bonding networks absent. However, where other spectroscopic techniques simply classify structure other than regular secondary structure as random coil, ROA has shown an extra incisiveness in being able to distinguish between different types of so-called random coil structure (Smyth et al., 2001). As a result of several recent ROA studies (Blanch et al., 2000; Smyth et al., 2001; Syme et al., 2002), many natively unfolded proteins or proteins with irregular folds are believed to contain the same, previously unassigned structural element. Far from being random coil, this is believed to be a much more open helical type structure known as poly-proline II (PPII) helix. PPII helix can adopt two helical conformations and in aqueous solution is a left-handed helix with 3 residues per turn (Woody, 1992). Recent studies have shown that PPII exists in significant proportions in globular proteins (Adzhubei et al., 1987; Adzhubei & Sternberg, 1993). Also, based upon proposals by Tiffany and Krimm (1972), VCD studies have shown PPII signatures (Dukor & Keiderling, 1991).

However, ROA has shown itself to be particularly sensitive to PPII helix, as shown in Chapters 6 and 7 of this thesis. Table 2.1. summarizes the $\phi$ and $\psi$ angle values for common secondary structure types.

**Table 2.1.** Range of $\phi,\psi$ angles for different types of secondary structure

| Structural Type | $\varphi$ angle | $\psi$ angle |
|---|---|---|
| α-helix | -60° | -45° |
| β-sheet (antiparallel) | -142° ±13° | 145° ± 13° |
| β-sheet (parallel) | -118° ± 13° | 112° ± 13° |
| PPII helix | -120° to -170° | 110° to 170° |
| β-turn (type I) | $\varphi_1 = -54°$ ; $\varphi_2 = -95°$ | $\psi_1 = -11°$ ; $\psi_2 = 9°$ |
| β-turn (type II) | $\varphi_1 = -48°$ ; $\varphi_2 = -88°$ | $\psi_1 = 137°$ ; $\psi_2 = -20°$ |
| β-turn (type III = $3_{10}$-helix) | $\varphi_1 = -45°$ ; $\varphi_2 = -48°$ | $\psi_1 = -36°$ ; $\psi_2 = -36°$ |

A visual representation is obtained by plotting the $\phi$ and $\psi$ angles for polypeptides against each other (Ramakrishnan & Ramachandran, 1965). This plot, known as a Ramachandran plot, is a convenient way to summarize the $\phi$ and $\psi$ angle distribution of amino acids within a protein and a general plot is illustrated in Fig 2.2.
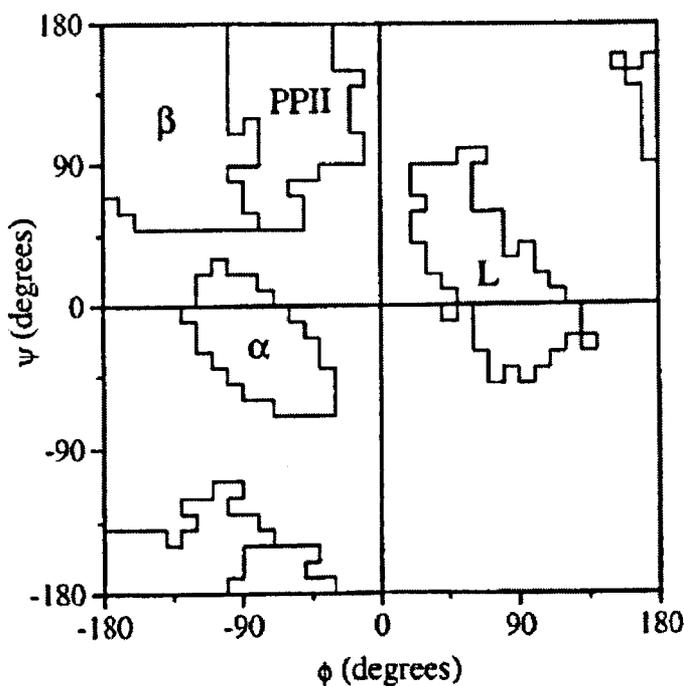


**Figure 2.2.** The Ramachandran plot shows the distribution of $\phi,\psi$ angles for a protein.

The amounts of secondary structure elements within a protein define the structural class of that protein. The strongest bands in the ROA spectrum of a protein are generated by secondary structure and as a result, ROA is proving to be a valuable tool in determining structural class. Compared with other spectroscopic data, protein ROA spectra contain the largest amount of information useful for determining structural class. Table 2.2. describes the criteria by which the structural class of a protein is defined.

**Table 2.2.** Definition of protein structural class reproduced from Kabsch and Sander (1983)

| Protein class | Structural criteria |
|---|---|
| alpha | $\alpha \geq 40\%,\ \beta \leq 5\%$ |
| beta | $\alpha \leq 5\%,\ \beta \geq 40\%$ |
| alpha + beta | $\alpha \geq 15\%,\ \beta \geq 15\%;\ > 60\%$ of $\beta$-strands antiparallel |
| alpha / beta | $\alpha \geq 15\%,\ \beta \geq 15\%;\ > 60\%$ of $\beta$-strands parallel |
| unordered | $\alpha \leq 10\%,\ \beta \leq 10\%$ |

**Structural motifs**

Certain arrangements of secondary structure elements are repeatedly found in proteins that define the basis of their tertiary structure. These arrangements are known as structural motifs, and have a significant role in determining protein ROA spectral patterns. As well as obtaining the tertiary fold of a protein from its ROA spectrum, more detailed information relating to individual structural motifs is also a realistic prospect.

The Greek key is a common motif and is formed when four $\beta$-strands align to form an *anti-parallel* $\beta$-sheet. The shape is similar to a design found in Greek pottery, and is illustrated in Figure 2.3A. The motif is not associated with any particular biological function, but is found in many proteins. The $\beta$-meander is very similar and is shown in Figure 2.3B. It is a simple motif consisting of *anti-parallel* $\beta$-sheets linked sequentially by short loops. Figure 2.3C shows an beta-alpha-beta ($\beta$–$\alpha$–$\beta$) unit, which is a very common way of connecting *parallel* $\beta$-strands.

**A - Greek key**    **B - Beta meander**    **C - Beta-alpha-beta unit**
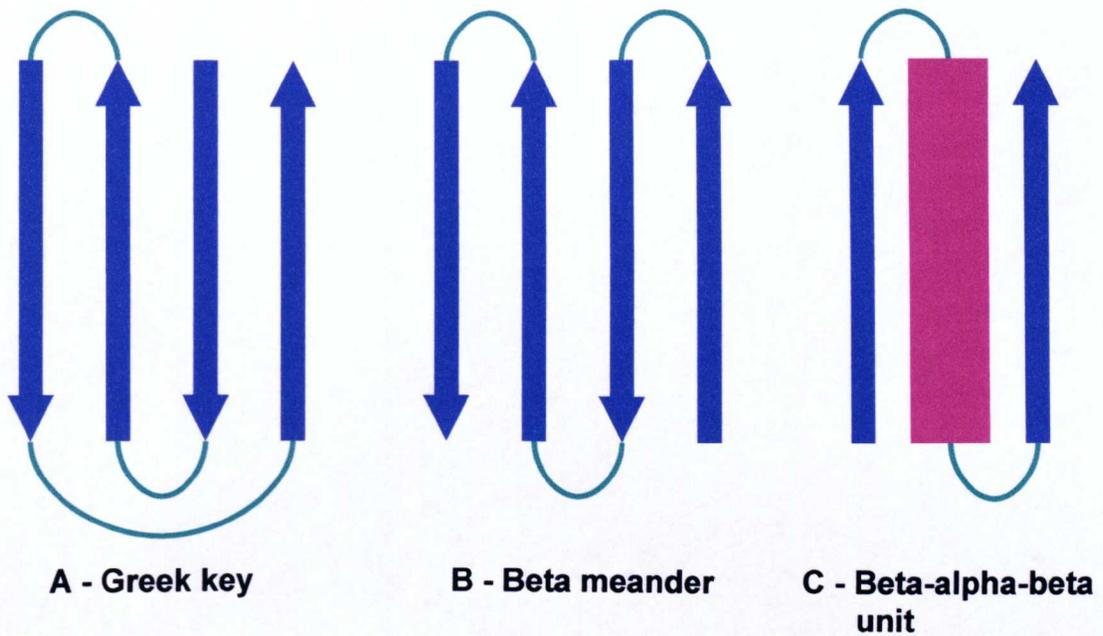
**Figure 2.3** Common structural motifs observed in many folded proteins. β-strands are represented by blue arrows, α-helices by purple blocks and loops by green connecting lines.

The helix packs against the β-sheets such that the hydrophobic face of the helix interacts with hydrophobic regions of the β-sheet. A common structural motif of particular interest in the context of this work is the jelly roll. A jelly roll results from two Greek key motifs that are linked by two connections and is shown in Figure 2.4.
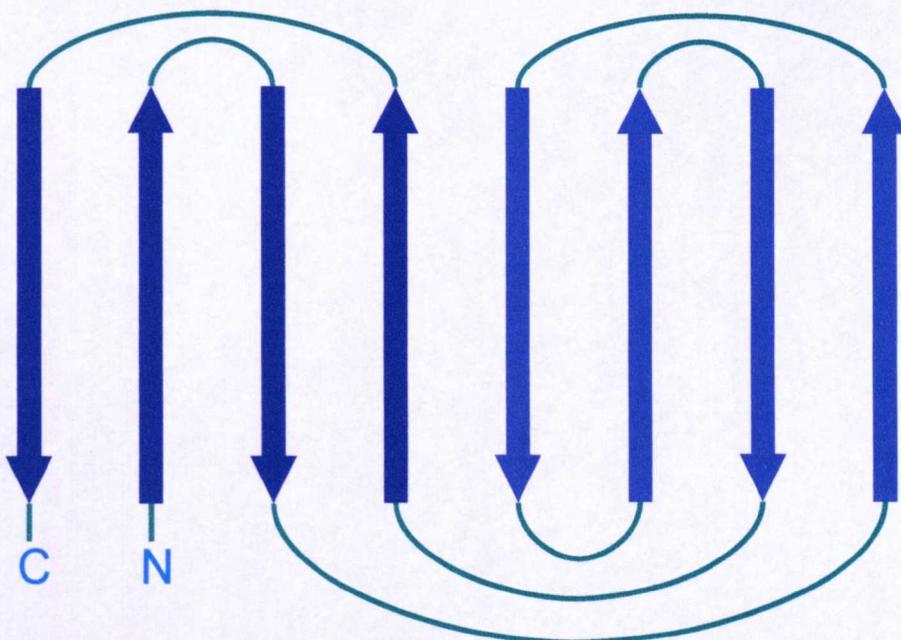


**Figure 2.4** The jelly roll motif is formed by the linkage of two Greek key motifs as shown. The jelly roll motif forms a β-barrel structure and is classified as one of the nine superfold classes.

The jelly roll motif forms a β-barrel and is the structural basis of the jelly roll superfold class illustrated in Figure 2.5. A number of examples of this superfold class have been studied by ROA spectroscopy and generate highly characteristic band patterns, making the jelly roll β-barrel an ideal model structure. The jelly roll motif is also found in the coat proteins of octahedral viruses which are of great current interest in ROA work.

**Tertiary structure**

The next level of the protein structure hierarchy is concerned with the spatial arrangement of secondary structure elements with respect to each other. Known as the tertiary structure of the protein, this level of structure is defined by the folding characteristics of the polypeptide chain, and is a result of the chemical properties of the amino acids and specifically the bulky side groups and how they hydrogen bond with each other.

The type of tertiary structure or 'fold' that a protein adopts therefore depends primarily upon the specific amino acid sequence of that protein, and hence is governed by the amount of the various secondary structure elements as well as loops and turns. Since ROA probes the chirality of the whole peptide backbone, it is sensitive to the presence of these sections of connecting structure as well as secondary structure. This means that the ROA spectrum of a protein reflects the three-dimensional fold of the protein as a whole.

Coupled with the fact that extraneous signals from side chains are minimized since ROA generation is a localized phenomenon centred around the chiral alpha carbons in the peptide backbone, ROA provides a uniquely incisive picture of the shape of the protein molecule. The proteins included in this study have been carefully selected across the whole range of structural classes, and incorporate at least 15 different fold types. The CATH classification system (Orengo et al., 1997) defines nine so-called superfolds (Orengo et al., 1994), which between them account for ~50% of all known fold types. Of these, six were available for study by ROA as part

of this work and are presented in subsequent chapters. Figure 2.5 illustrates the nine superfold classes.
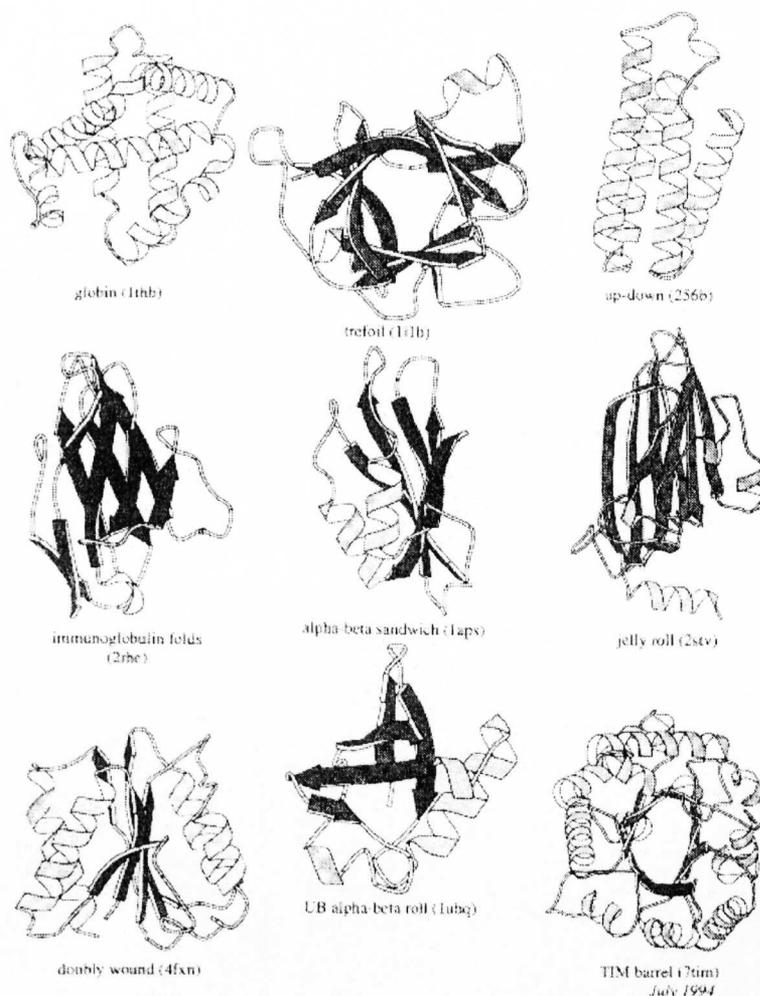


**Figure 2.5** The nine protein superfold classes as defined by Thornton et al. (reproduced from *Phil. Trans. R. Soc. Lond.* B, **348** , 71-79 (1995)).

## 2.3 **Vibrational modes of proteins**

ROA is a form of vibrational spectroscopy, and as such yields detailed and highly informative data. Since an atom in a molecule has three degrees of freedom, a molecule with N atoms has 3N degrees of freedom. Subtracting the modes that represent rotation and translation of the molecule, there are 3N - 6 normal modes of vibration in a molecule (Wilson et al., 1955). It is clear to see that in molecules the

size of typical proteins, normal mode analysis is extremely (if not impossibly) complex.

However, the principal modes of vibration of key interest in protein ROA work are the amide I and amide III normal vibrational modes. These modes are depicted in Fig. 2.6., and their chief components summarized in Table 2.3.
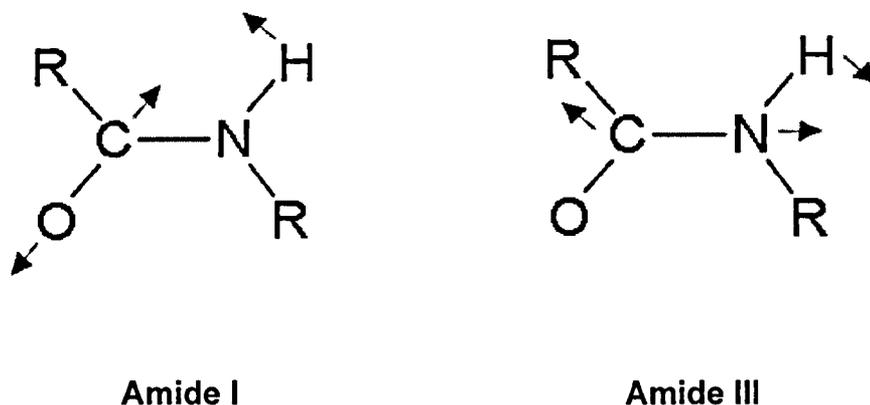
**Amide I**                              **Amide III**

**Figure 2.6.** Amide I and Amide III normal vibrational modes of proteins.

**Table 2.3.** Summary of key vibration components in normal vibrational modes of proteins (Diem, 1993).

| Vibrational mode | Vibrational components | Wavenumber range |
|---|---|---|
| Amide I | 80% CO stretch | ~1630 - 1770 cm$^{-1}$ |
| Amide II | 60% NH in-plane bend; 40% CN stretch | ~1480 - 1560 cm$^{-1}$ |
| Extended Amide III[*] | Various combinations of CN stretch, NH in-plane bend and C$^\alpha$ deformations | ~1220 - 1360 cm$^{-1}$ |

[*] An extension of the classic amide III region defined as ~1240 – 1300 cm$^{-1}$ (Diem, 1993)

## 2.3  The ROA spectrum of a protein

The current experimental configuration allows the measurement of the ROA spectra of proteins from ~600 - 1770 cm$^{-1}$ digitized at intervals of ~3 cm$^{-1}$. This is the

typical range of the spectra presented in this thesis. The main regions of interest in protein ROA spectra correspond to the amide vibrational modes described in Table 2.3 and are illustrated in Figure 2.7. This spectrum of bovine α-lactalbumin was recorded over a period of around 6 hours in aqueous solution at room temperature.

The Raman scattering event is around 3 orders of magnitude faster than the resulting ROA signal that represents a superposition of the ROA spectra of each conformational state explored by the protein in its native solution structure.
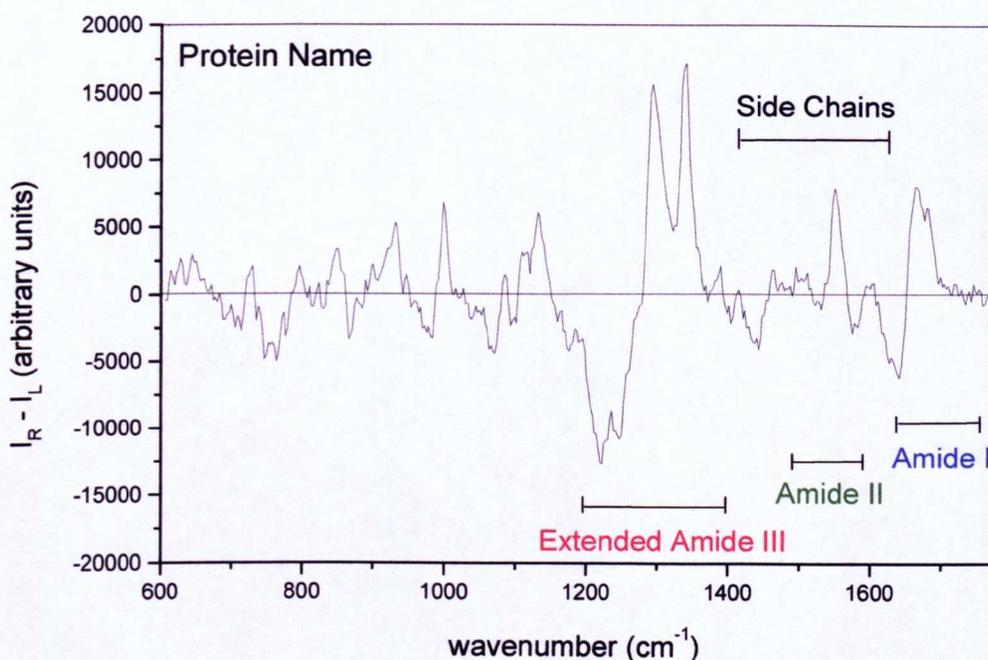


**Figure 2.7.** Template of a protein ROA spectrum denoting the major features of interest. Note that ROA spectra are bisignate, i.e. they contain both positive and negative peaks.

## 2.5 The analysis of protein ROA spectra

With a lack of theoretical models, the relationship between protein structure and ROA spectral features has been built up on an empirical basis, mainly through observed band patterns in the amide I and amide III regions. By correlating these bands to known structural features from X-ray crystal structures and NMR solution structures, it has been possible to construct a table of current ROA band assignments. The problem of identifying relationships between observed data and known structural details of the analytes is an example of the inverse spectral problem. In order to help

solidify the more important band assignments, it has been useful to consider model structures such as helical polypeptides and proteins with a predominance of one type of secondary structure element, examples of which will be discussed in subsequent chapters.

A source of important information with regards to the construction of a useful band assignment table has been to study the effects of solvent on the structure of proteins. By using $D_2O$ instead of $H_2O$ and comparing the difference between the normal and deuterated ROA spectra, it is possible to investigate which bands are coming from solvent exposed structures. This method is generally not useful for comparisons between different samples as proton exchange rates differ greatly, even within different regions of the same protein. Therefore, information from such experiments can act merely as a guide to the nature of solvent interaction/hydration and is by no means an exact science. However, one of the best examples of differences between the ROA spectra of a protein in both $H_2O$ and $D_2O$ has been the assignment of hydrophilic and hydrophobic α-helical bands, as well as helices of an amphipathic nature.

The analysis of protein ROA spectra is a multi-faceted task that relies heavily upon the detailed visual inspection of the data and comparison of spectral features to previously observed features that have been assigned physical significance. Previous work has shown that the main secondary structure elements and structural motifs generate ROA bands consistently in the same spectral regions. However, the current goal in the analysis of protein ROA spectra has been to take advantage of the unique ability of ROA to see loop and turn structure, as well as random coil and PPII helix, to give a picture of the protein molecule as a whole (specifically the tertiary fold) that is unrivalled by other conventional spectroscopic techniques. However, as the spectra presented in the following chapters will reveal, the complexity and richness of detail within protein ROA spectra poses a great challenge to analyst. The numerous individual structural elements and structural motifs within proteins in general allow a very large amount of variation, even within the bounds of known tertiary fold types.

Table 2.4. serves as a useful guide to interpreting ROA spectra by visual inspection. All ROA band values are in wavenumbers (cm$^{-1}$) and additional footnotes apply to notable individual exceptions to general observations.

**Table 2.4.** Current protein ROA band descriptions and assignments.

| Band Range | Intensity* | Sign | Assignment (Likelihood)† |
|---|---|---|---|
| 870 – 950 | m | + | α-helix (h) |
| 1000 – 1050 | m | + | β-strand (m) |
| 1080 – 1130 | m | -/+ | α-helix (m)[a] |
| 1215 – 1225 | s | - | β-strand (h) *hydrated?* |
| 1238 – 1250 | s | - | β-strand (h) |
| 1265 – 1295 | s | + | loops and turns (m) |
| 1295 – 1312 | s | + | hydrophobic α-helix (h)<br>$3_{10}$-helix (m)<br>residues in loops in α/$3_{10}$ region (m) |
| 1338 – 1346 | s | + | α-helix, hydrated (h) |
| 1340 – 1380 | m | - | β-hairpins (h) |
| 1400 – 1480 | w | complex patterns | side chains (aromatic at lower, aliphatic at higher cm$^{-1}$) (h) |
| 1530 – 1580 | s | +/- | tryptophan side chains (h) |
| 1600 – 1630 | m | various patterns | aromatic side chains (h) |
| 1630 – 1665 | m | -/+ | α-helix (h)[b] |
| 1645 – 1680 | m | -/+ | β-sheet (h) |
| 1670 – 1685 | m | + | disordered (PPII?) (h) |

[a] - not in polyalanine nor AK21 (side chain dependence?) [b] – not in polyalanine
*- w, m and s denote weak, medium and strong respectively
† - (l), (m) and (h) denote low, medium and high likelihood respectively

However, it is evident that proteins that are known to be structurally similar do show the same underlying ROA band patterns, with differences (subtle or less so) in

fine structure that can be easily assigned to real differences in structure in a consistent fashion. Ideally, it should be possible to deconstruct the ROA spectrum by deconvolution to see what contribution each type of structure is making to the overall structure. However, the problems associated with this task are numerous. The problem of spectral interpretation and factors contributing to protein structure are summarized in Figure 2.8. A major problem in forming a relationship between real protein structure and ROA spectral features is the fact that for many of the proteins now being

| Experiment | | Analysis | | Correlation | |
|---|---|---|---|---|---|
| **PROTEIN** | ⟶ | **SPECTRUM** | ⟶ | **CONCLUSION** | ⟶ **PROTEIN** |

| Structure (real) | ROA band pattern | Structure (abstract) | Structure (real) |
|---|---|---|---|
| Experimental Conditions | Positions (cm$^{-1}$) | Hydration | |
| Behaviour | Sign (+/-) | | |
| | Strength | | |
| | Width | | |

Components of protein

| **Helix content** | **Loops and turns** | **Structural motifs** | **Macro structures** |
|---|---|---|---|
| Alpha- | Number | Number | Fold type |
| 3$_{10}$- | Length of loops | Connectivity | No. of domains |
| PPII- | Relative amount | Architecture | Oligomerization |
| | | | Aggregation |
| **Strand content** | **Side chains** | **Stabilizing features** | SS Interactions |
| Number | Type | Disulphide links | **Solvent effects** |
| Length | Number | Metal ions | |
| Sheet types | Orientation | Co-operative folding | Amphipathicity |
| Topology | | Side chain | Hydrophobicity |
| Connectivity | | interactions | Hydration |

**Figure 2.8** Flow diagram (top) describing the process of interpretation of protein ROA spectra. Structural factors (bottom) within proteins all affect the appearance of the band patterns in the ROA spectra of proteins, and must somehow be incorporated into the analysis step to obtain an abstract structure from the ROA spectrum that is related to the real structure of the protein.

studied by ROA, X-ray crystal structures are not available and there is no other structural information to provide a basis for deconvolution. Chapter 8 describes the

application of a pattern recognition technique, principal component analysis (PCA), for the analysis of protein ROA spectra which does not involve any such pre-treatment of the original data, yet is able to distinguish the structural class of the each protein in the data set in a self-consistent manner, with a high degree of success.

## 2.6  Summary

The complexity of protein structure has meant that conventional spectroscopic techniques have struggled to offer much more beyond approximate secondary structure content. In addition, the bulk of protein molecules is built up from side chains branching from the polypeptide backbone. Since ROA directly probes the chirality within the polypeptide backbone, the spectra contain a wealth of information about the backbone conformation which is unrivalled by other conventional techniques. This wealth of information must be correlated with known structural information in order to form a basis for the analysis of protein ROA data. This has been achieved to a great extent by the detailed visual inspection of spectra of as many types of proteins with as diverse a range of structural properties as possible. This 'survey' of protein structure is on-going work, and the spectra presented in the following chapters represent the latest data set, on which new analytical techniques have been tried and tested.

The most apparent conclusion that can be drawn from the current protein ROA data set is that there is a definite relationship between the variety of structural elements that are known to be present in the proteins and the bands which appear in their ROA spectra. Furthermore, the overall ROA band patterns are visually similar between proteins of the same fold. The complexity of the spectra is to be expected given the range of structural features contributing to the overall shape of the protein, and this has led to significant problems in the qualitative comparison and analysis of protein ROA spectra. The approach of attempting to deconvolute ROA spectra in order to ascertain which spectral features are related to which structural features has been largely abandoned in favour of the more mathematically rigorous method of principal component analysis, which is discussed in Chapter 8.

In order to solidify band assignments and further understand the relationship between protein structure and protein ROA band generation, more and more proteins with different structural properties are to be studied. Due to the number of proteins already studied and the wealth of information they have yielded, it is now becoming straightforward to determine structural class, approximate secondary structure content and the tertiary fold of a protein from its ROA spectrum alone.

# Chapter Three

# Alpha Proteins

## 3.1 Introduction

One of the central problems encountered in the analysis of ROA spectra is the availability or otherwise of model structures to solidify band assignments. Fortunately there are many α-helical proteins and model α-helical homopolypeptides, and these have been of key importance in the interpretation of protein ROA spectra. The term 'alpha proteins' is used in a more general sense than in the literature and for the purposes of this discussion includes both all alpha (~35-60% α-helix) and mainly alpha (~15-35% α-helix). As well as proteins that only have α-helix as their major component, the term 'mainly alpha' also encompasses a wide range of proteins that contain a significant amount (up to ~15%) β-strand. At this point, it is useful to refer to the modified structural class definitions (Table 2.2) in the previous chapter. An important distinction should be made between mainly alpha proteins and 'alpha beta' proteins. The latter are the subject of a later chapter. In accordance with the CATH classification system, both contain comparable amounts of both α-helix and β-strand.

## 3.2 Experimental

A sample of human serum albumin (HSA) was obtained from ICN Pharmaceuticals. Bovine serum albumin (BSA), hen lysozyme and bovine α-lactalbumin was obtained from the Sigma Aldrich Company. Samples of human and equine lysozyme were supplied by L. Morovoza-Roche of the University of Oxford. Bovine pancreatic insulin was obtained from the Sigma Aldrich Company and was prepared at pH 1 to obtain the monomeric solution. Dr. G. Tennent of the Royal Free University College Medical School supplied a sample of creatine kinase. Also, the filamentous bacteriophage strains IKe, Pf1 and M13 (supplied by L. A. Day of the Public Health Research Institute, New York) will be included in the discussion briefly, although these spectra were not obtained by the author and as such will not be presented in this chapter.

## 3.3  Results and discussion

### Serum albumins

The serum albumins are the main component of blood and are responsible for ligand transport and maintaining blood pH (Carter et al., 1994). Bovine serum albumin (BSA) and Human serum albumin (HSA) are both large proteins with 585 amino acid residues and a molecular weight of ~65 kDa. The serum albumins are highly structured proteins containing 17 disulphide bonds and 67% α-helix. Apart from some viral coat proteins, these proteins contain the highest number of α-helices of any studied by ROA and as such are extremely valuable in understanding the nature of hydration effects upon helices (Wilson, 1996). In addition to the highly helical character of these proteins, there is also a great number of connecting loops and turns that hold the helices in a stable tertiary fold. The serum albumin fold type as defined by the CATH classification system is similar to a mixture of the globin and up-down superfold classes as illustrated in Figure 2.3., in that the main body of the protein consists of a mass of inter-connected α-helices with a range of relative spatial alignments and interfacial boundaries.
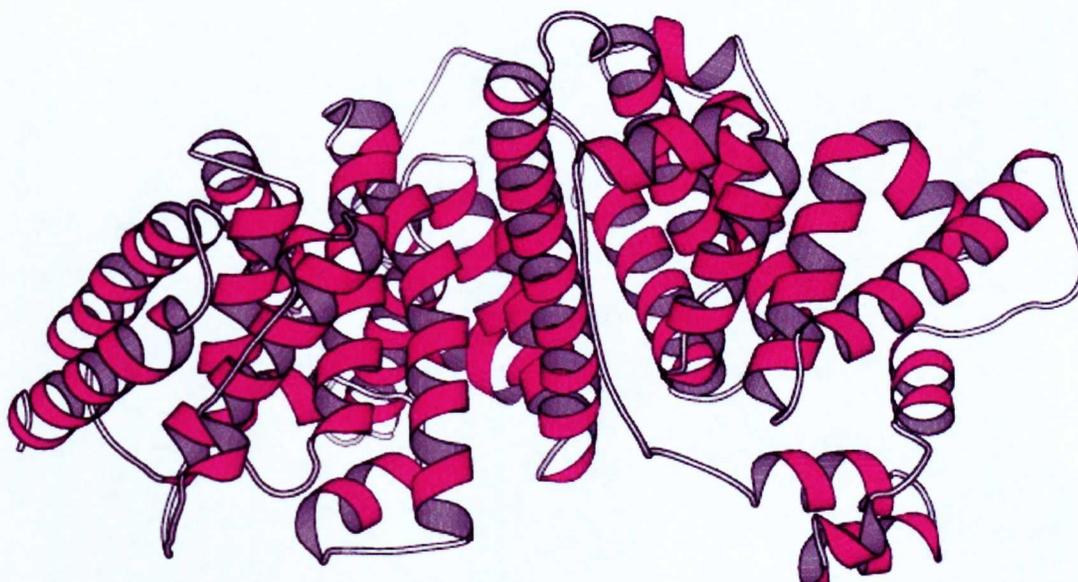


**Figure 3.1.** The crystal structure of human serum albumin (HSA), pdb code 1bj5. The structure of bovine serum albumin (BSA) assumed to be almost identical.

The crystal structure of HSA is available (PDB code 1bj5, Curry et al., 1998) and is shown in Figure 3.1. The crystal structure of BSA is not available but the primary structure is homologous to ~98% of that of HSA, and therefore the structure is assumed to be almost identical to that depicted in Figure 3.1. The backscattered Raman and ROA spectra of HSA and BSA are shown in Figure 3.2
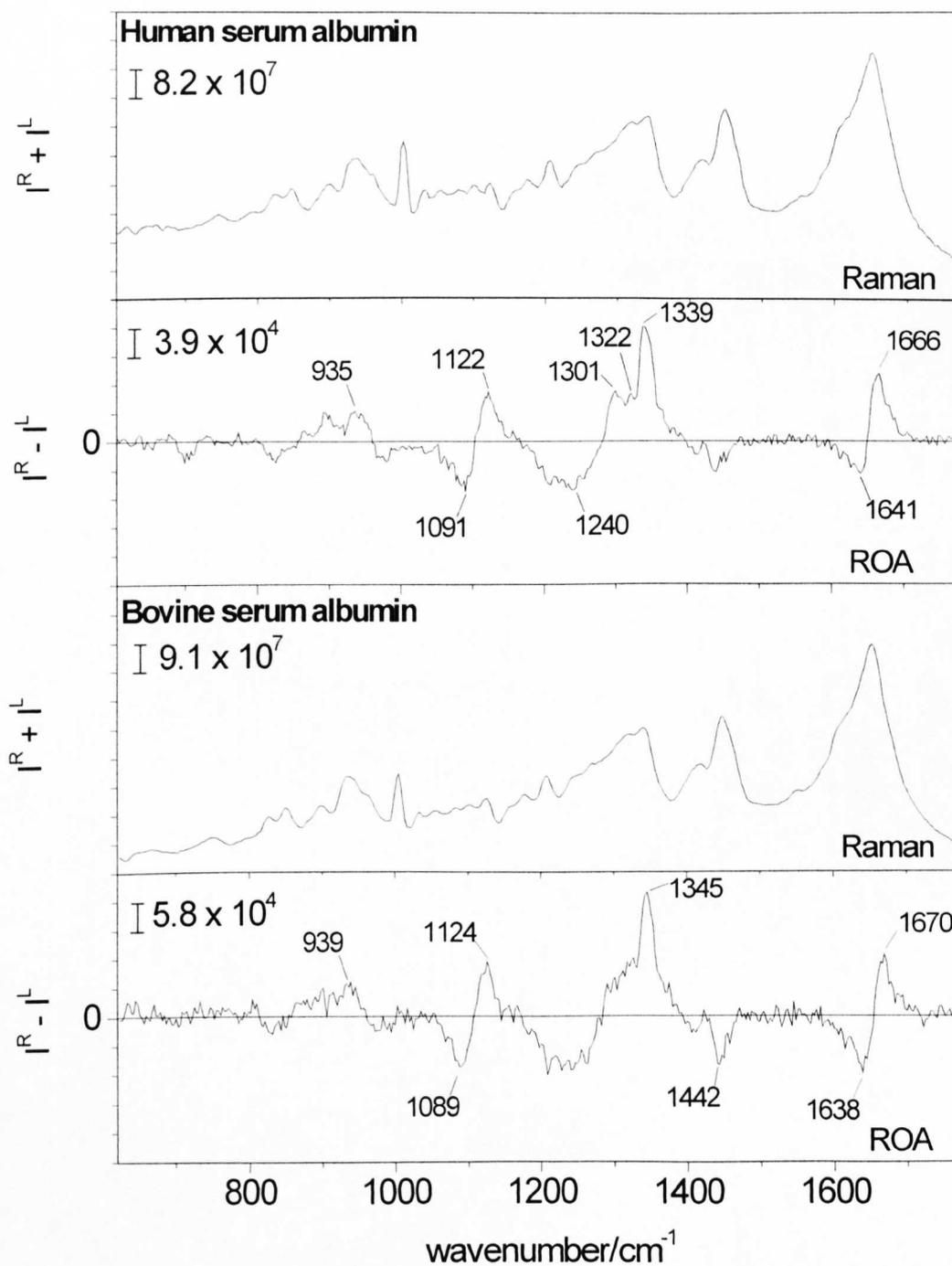


**Figure 3.2**. The backscattered Raman and ROA spectra of HSA (upper) and BSA (lower). Both spectra were measured at 100 mg/mL in 0.1M acetate buffer at pH 5.4.

In addition to the spectra shown in Figure 3.2., the ROA spectrum of HSA was also collected in $D_2O$. The resulting spectrum was reasonably informative but the data was not of sufficiently superior quality to that collected previously (Wilson, 1996) to merit further discussion. It is sufficient to say that the main features of the spectrum were reproduced by this current study, but that no significant refinement to previous work could be made on the basis of the new data.

It is not surprising that the spectra shown in Figure 3.2. are extremely similar as the proteins themselves are assumed to be almost identical. The actual samples were obtained from different suppliers and originate from different species, and the difference in the purity of the samples is not accurately known. However, despite these differences, it is clear that ROA is able to distinguish the fact that the structures of these two proteins are very similar. Each spectrum is dominated by $\alpha$-helical signature bands, primarily at ~1340 $cm^{-1}$ indicating hydrated $\alpha$-helix and a sharp positive band at ~1663 $cm^{-1}$ in the amide I region indicating helical structure. There is also a positive peak at ~1300 $cm^{-1}$ indicative of hydrophobic $\alpha$-helix, as well as a couplet with a positive peak at ~1120 $cm^{-1}$ commonly observed in proteins with a significant proportion of $\alpha$-helix. The spectrum of HSA obtained in $D_2O$ revealed the disappearance of the strong sharp band at ~1340 $cm^{-1}$ in the amide III region, similar to the finding of the previous study by Wilson et al. (1996), strongly suggesting that the exchange of amide protons from solvent exposed $\alpha$-helices was taking place, and hence that the 1340 $cm^{-1}$ band assignment is largely due to hydrated $\alpha$-helix.

Further evidence of the large proportion of $\alpha$-helix present comes from the skeletal backbone region where positive structure peaking at ~892 $cm^{-1}$ and ~934 $cm^{-1}$ is seen. Again, this is a common feature of proteins with significant helical character. The negative band at ~1240 $cm^{-1}$ is interesting as it seems to suggest that these proteins contain some beta character, although the lack of any negative structure at ~1340 - 1380 $cm^{-1}$, coupled with the low wavenumber couplet in the amide I region suggests that this beta character may be limited to short segments of isolated $\beta$-strand, as opposed to $\beta$-sheet. The significant loop and turn structure may also contribute to the negative band centred at ~1240 $cm^{-1}$ and probably accounts for the positive peak between the two main $\alpha$-helical bands at ~1318 $cm^{-1}$.

## The lysozyme fold family

Possibly the most interesting set of samples to be studied by ROA spectroscopy, the lysozyme fold family provides an excellent opportunity to study a set of proteins with the same known tertiary fold. Due to the abundance of information available on the structure of one member of this large group of proteins, hen lysozyme, this protein is routinely used as the standard by which ROA spectra are measured. The structural richness and physical stability of hen lysozyme have made this protein a perfect example of how much and what type of structural information can be obtained from a protein ROA spectrum. The current group of lysozymes include hen, human and equine lysozyme together with bovine α-lactalbumin which has a similar X-ray crystal structure, and their ROA spectra are presented in this section. With the exception of equine lysozyme which is included for completeness and as part of a more general study, these spectra were obtained by the author, although all have been subject to prior and/or subsequent detailed investigations (Blanch et al., 2000a, 2000b; Barron et al., 2000).

Lysozymes act as bacteriolytic enzymes by hydrolyzing the beta bonds between N-acetylglucosamine and N-acetylmuramic acid in the peptidoglycan of prokaryotic cell walls. α-lactalbumin is a milk protein which has a role in regulating the subunit of lactose synthetase. α-lactalbumin changes the substrate specificity of galactosyltransferase from N-acetylglucosamine to glucose in the mammary gland (de Gruyter, 1988). Crystal structures have been resolved for all four of these proteins, and are illustrated as Molscript diagrams in Figure 3.3. (Kurinov et al., 1995; Harata et al., 1998; Tsuge et al., 1992; Acharya et al., 1989). Lysozymes consist of two separate domains, an alpha domain and a beta domain (consisting of a single two-stranded β–sheet). These proteins contain between ~35-40 % α-helix and ~10-15% β-strand, qualifying them as mainly alpha proteins. The crystal structures of these proteins reveals the fact that both hydrophobic and hydrophilic α-helix is present.. These proteins are also known to contain between four and six tryptophan residues each that figure prominently in the ROA spectra of these proteins, and have been the subject of a recent ROA study (Blanch et al., 2001).
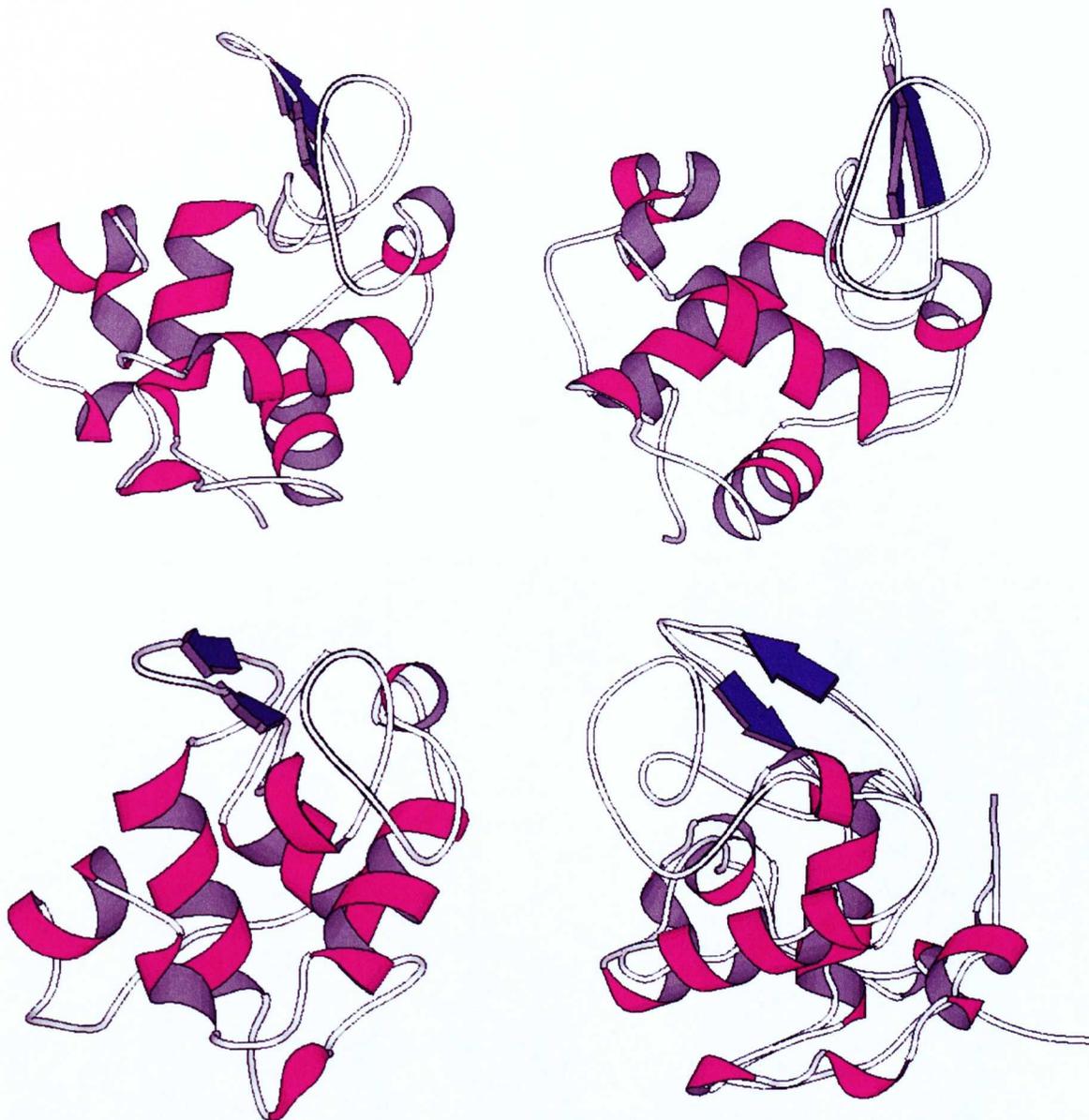
**Figure 3.3.** Molscript diagrams representing the crystal structures of four members of the lysozyme fold family. Clockwise from top – hen lysozyme (pdb code 1lse); human lysozyme (pdb code 1jsf); equine lysozyme (pdb code 2eql); bovine α-lactalbumin (pdb code 1alc).

The backscattered Raman and ROA spectra of these proteins are presented across Figures 3.4A and 3.4B. An additional representation (Figure 3.5) shows the normalized spectra of all four proteins on one page, allowing the similarity of the spectra to be more easily appreciated. This serves as a good example of the difficulties posed in attempting to compare several protein ROA spectra simultaneously and highlights the requirement for the development of a more rigorous method for this purpose, and is the subject of further discussion in a later chapter.
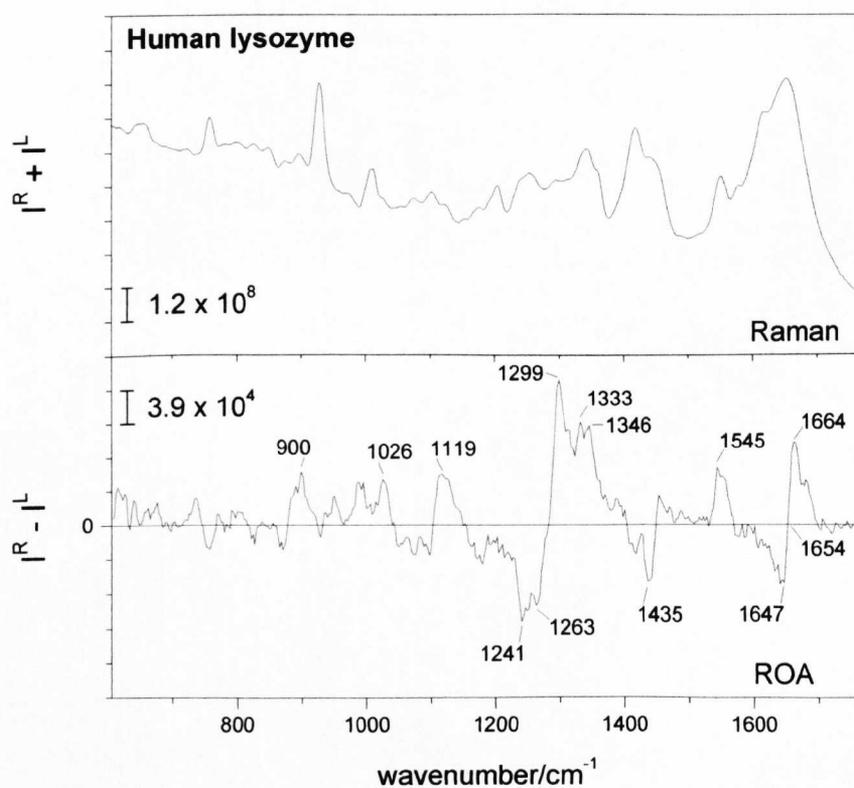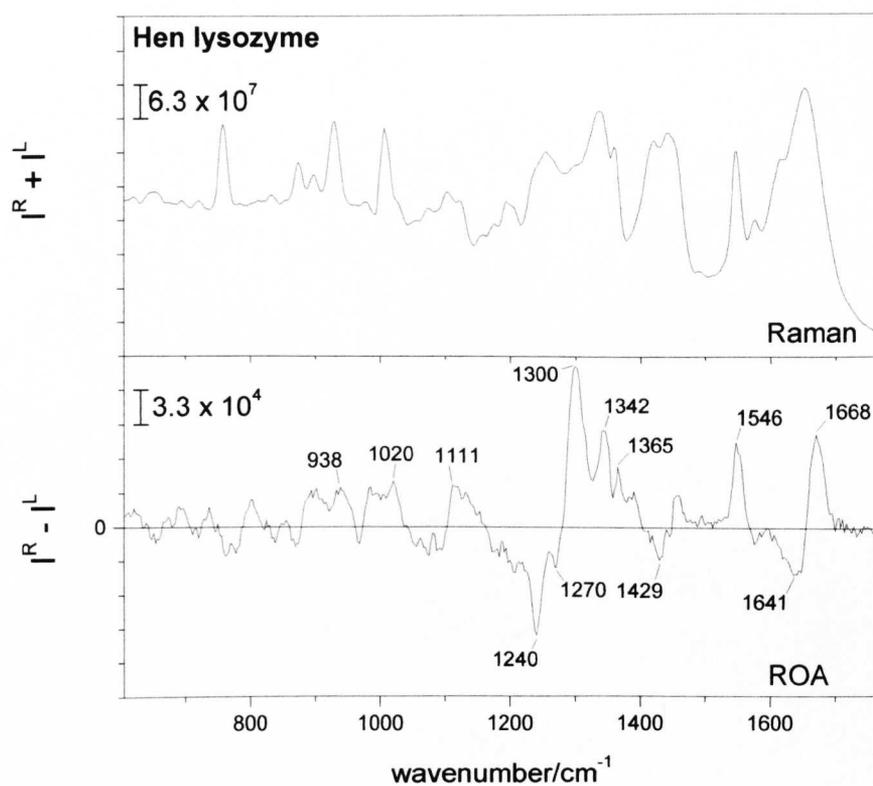
**Figure 3.4A** Top – Backscattered Raman and ROA spectra on hen lysozyme. Spectra measured at ~100 mg/mL in 0.1M acetate buffer at pH 5.4. Bottom – Backscattered Raman and ROA spectra of human lysozyme. Spectra measured at ~67 mg/mL in water at pH 7.
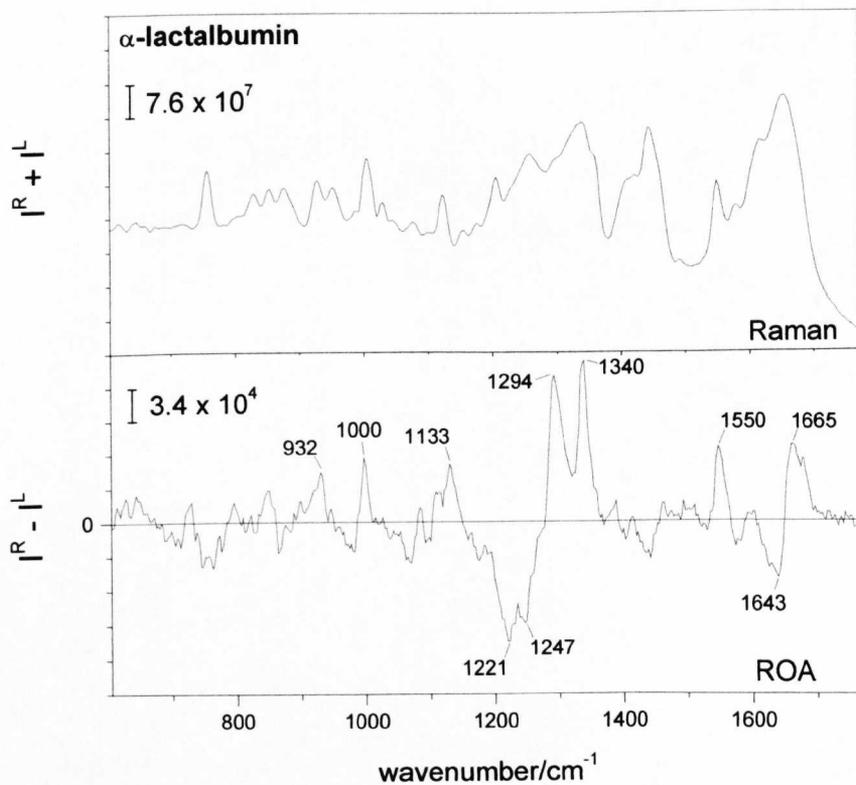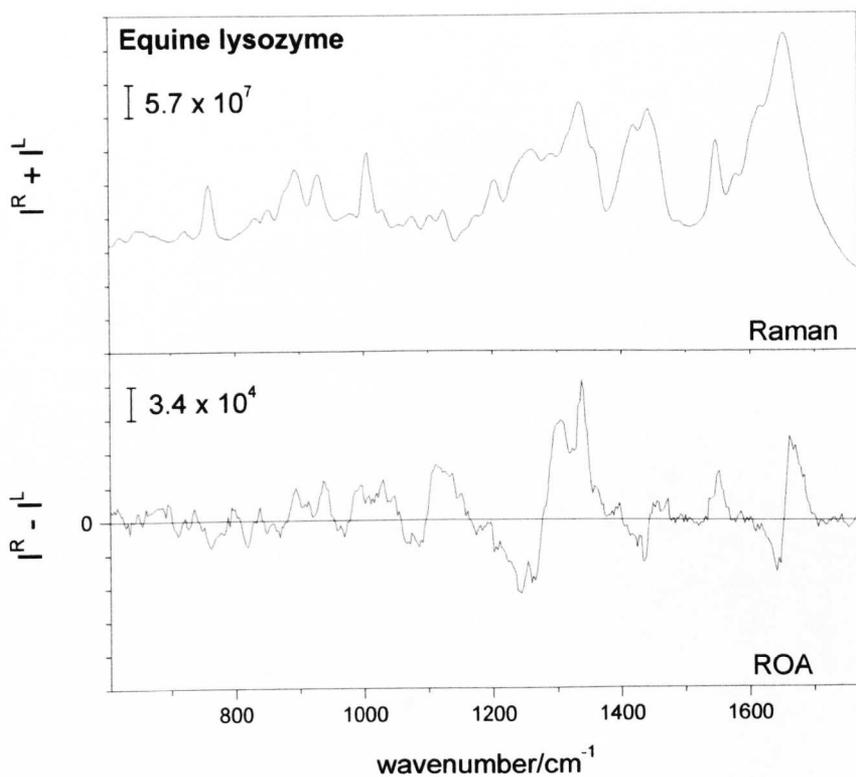
**Figure 3.4B** Top – Backscattered Raman and ROA spectra of equine lysozyme. Spectra were measured at ~100 mg/mL in water at pH 7. Bottom – Backscattered Raman and ROA spectra of bovine α-lactalbumin. Spectra were measured at ~100 mg/mL in 0.1M acetate buffer at pH 5.4.
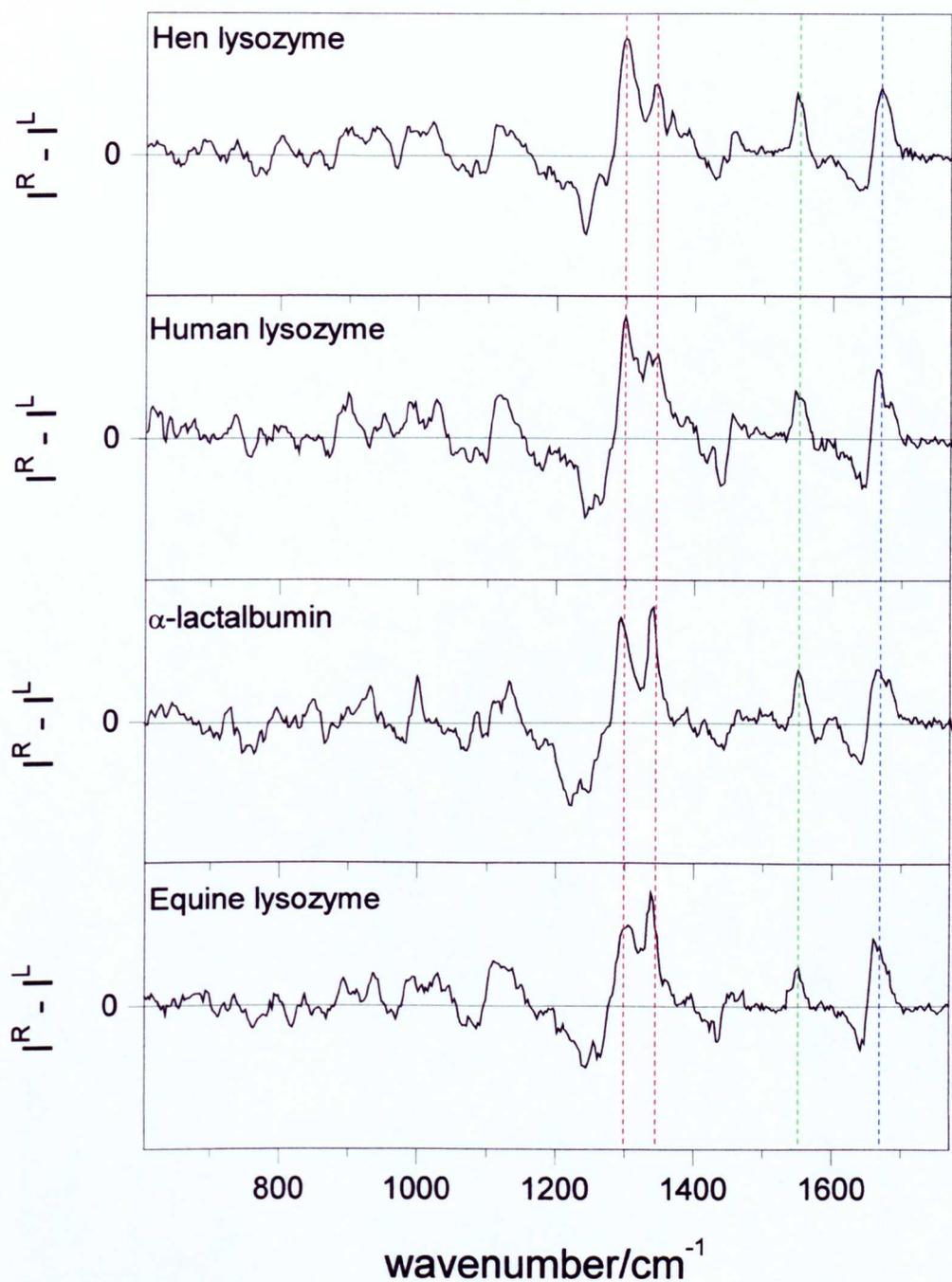
**Figure 3.5.** The backscattered ROA spectra of four members of the lysozyme family. Top to Bottom – Hen lysozyme, human lysozyme, α-lactalbumin and equine lysozyme. Common bands are highlighted for clarity. The amide III peaks at ~1300 and ~1340 cm-1 are shown in red, the amide I peak at ~1667 cm-1 is shown in blue. In addition, the positive band at ~1550 cm-1 originating from the tryptophan residues are shown in green.

Each of these four spectra share similar detail resulting in a similar overall ROA spectral band pattern, reflecting the fact that these proteins belong to the same fold class. The most prominent features are illustrated in an average spectrum shown in Figure 3.6, obtained by normalizing each of the four spectra shown in Figure 3.5 and adding them together. The positive peaks in the amide III region at ~1300 and ~1340 cm$^{-1}$ are due to hydrophobic and hydrated $\alpha$-helix respectively, with evidence for the later assignment based upon previous studies with $D_2O$ (Wilson, 1996) At higher wavenumber, the spectra show a great deal of similarity with a negative positive couplet from ~1420-1450 cm$^{-1}$ coming from tryptophan and contributions from $CH_2$ and $CH_3$ deformations (Wilson, 1996) and a strong positive ROA band at ~1555 cm$^{-1}$ also coming from tryptophan. The amide I region is similar in each spectrum, with a negative component at ~1640 cm$^{-1}$ and positive at ~1665 cm$^{-1}$, typical of mainly alpha proteins (tending to lower wavenumber than beta proteins). To the lower wavenumber side of the amide III region, there is also a great deal of similarity, with a strong negative band at ~1240 cm$^{-1}$ attributed to the lysozymes $\beta$-strand content. There is also positive structure centred at ~1120 cm$^{-1}$ which is observed in many mixed alpha and beta proteins. Characteristic $\alpha$-helix bands from ~900 – 960 cm$^{-1}$ (Wilson, 1996) are also seen in all four spectra, although the clearest examples are those of hen and equine lysozyme.
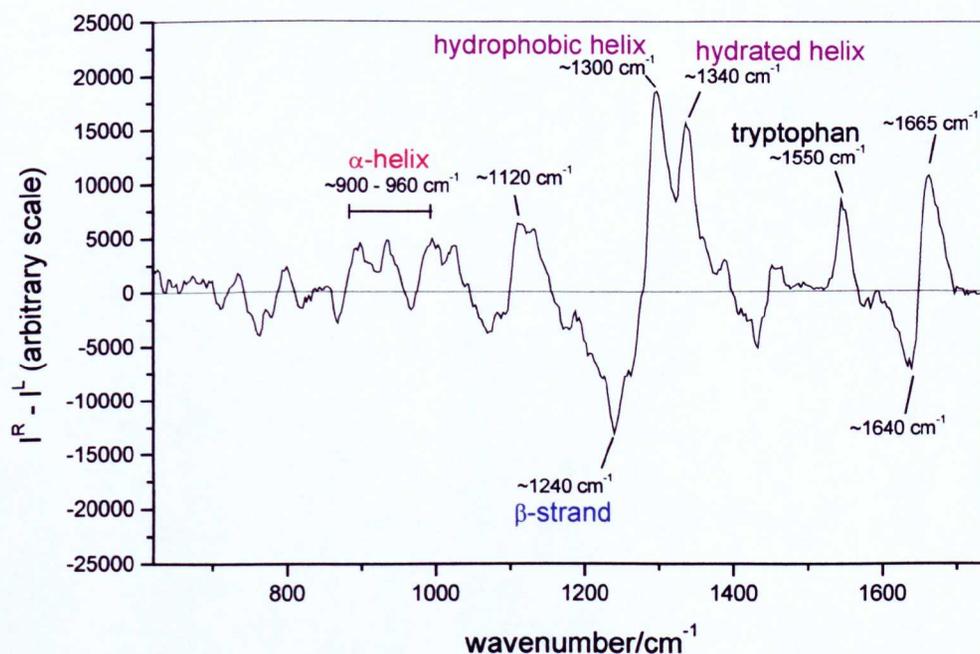


**Figure 3.6.** The average normalized ROA spectrum of a lysozyme. The main bands of interest are marked for clarity.

**Bovine insulin**

Insulin from bovine pancreas is a hormone with a molecular mass of 6kDa. Insulin usually exists in the native state as a dimer, but does form higher oligomers. However, it is the monomeric form that is of interest here and in order to achieve this state, it is necessary to go to low pH (~pH 1.0).
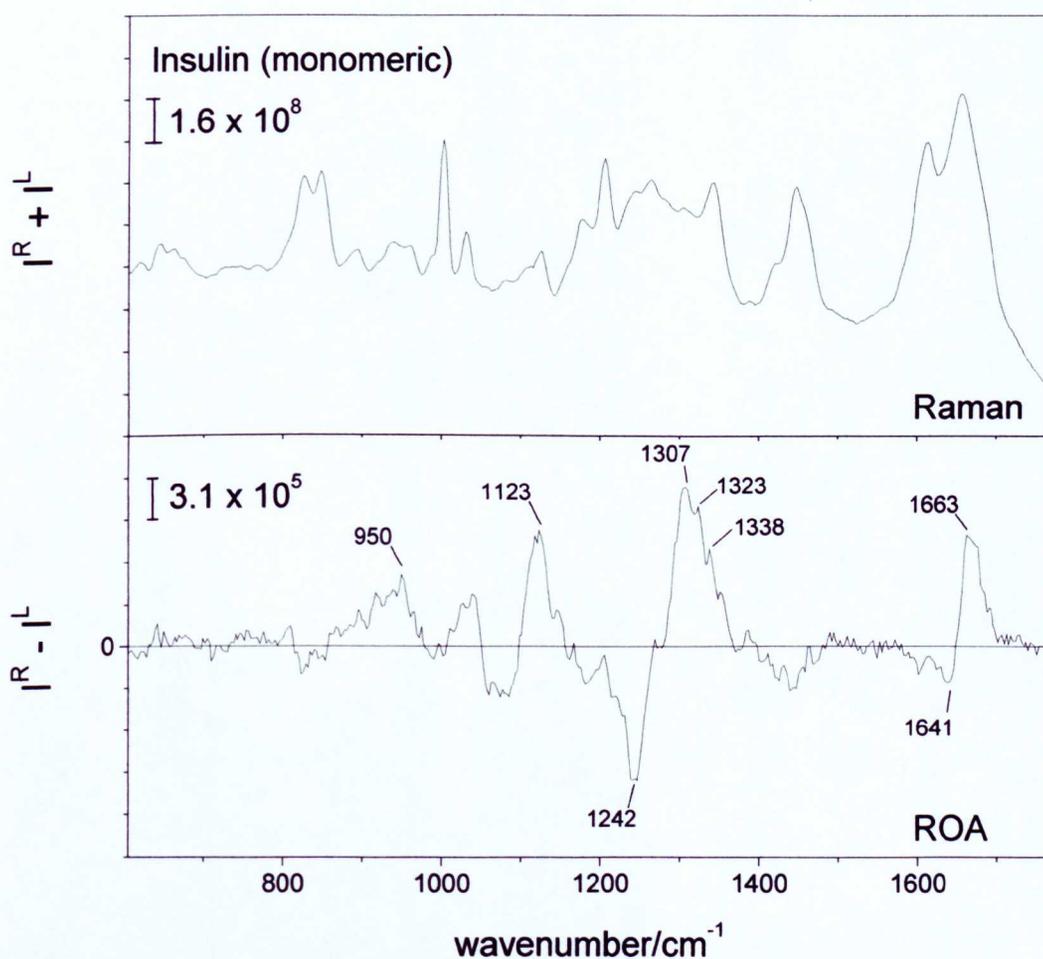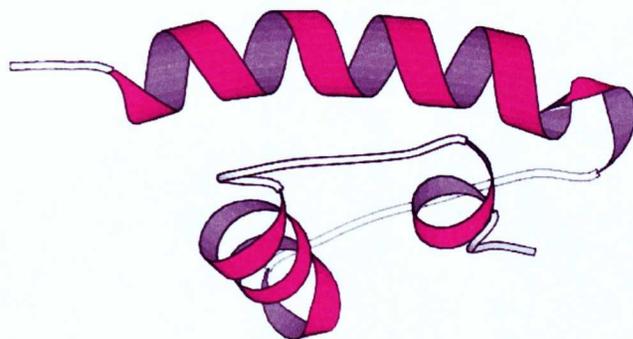


**Figure 3.7.** Molscript diagram and the backscattered Raman and ROA spectra of bovine insulin in the monomeric state. Spectra were measured at ~67 mg/mL in 0.1M acetate buffer with 1M NaOH added to ~pH 1.0.

Dynamic light scattering (DLS) was performed on the low pH insulin solution to confirm that it was in the monomeric state. The crystal structure of bovine insulin is available (Whittingham et al., 1998) and one domain of the protein is illustrated in a MOLSCRIPT diagram in Figure 3.7. Although classified as a mainly alpha protein, insulin contains some β-strand (not illustrated) which forms double-stranded sheet in the dimer.

The backscattered Raman and ROA spectra of monomeric insulin is shown in Figure 3.7. The ROA spectrum shares several similar features to the spectra discussed previously with a noticeably similar overall band pattern. However, where the lysozyme fold family have a clear positive band at ~1340 cm$^{-1}$, insulin has only a small peak at ~ 1338 cm$^{-1}$, suggesting that most of the α-helix is hydrophobic and that the fold structure of insulin is more open than that of the lysozymes. Evidence for β-strand can been seen with a very strong negative band at ~ 1242 cm$^{-1}$, but the amide I region suggests that this is in the form of unassociated β-strand as opposed to β-sheet as the negative positive couplet from ~1641-1663 cm$^{-1}$ is too low for β-sheet and similar to that seen in other mainly alpha proteins.

**Creatine kinase**

Creatine kinase catalyzes the reaction that converts ATP to ADP and may represent up to 10-20% of muscle cytoplasmic protein (Watts et al., 1973). The creatine kinase enzyme exists as a dimer with its subunits combining with others from different species to form active hybrids (Dawson et al., 1967). The monomer comprises two domains, a mainly alpha domain and an αβ 2-layer sandwich. The crystal structure of creatine kinase from rabbit muscle is available (PDB code 2crk, Rao et al., 1998) and is illustrated in Figure 3.8. as a MOLSCRIPT diagram. Fold characteristics of this protein are distinct from others shown in this chapter, although the alpha and beta regions are separated as in the lysozymes.

The backscattered Raman and ROA spectra are shown in Figure 3.8. The most significant features of the ROA spectrum suggest a mainly alpha protein although the strongest positive band in the amide III region peaks at ~1310 cm$^{-1}$, which is high for α-helix and suggests that there is a significant amount of loops and turns contributing

43
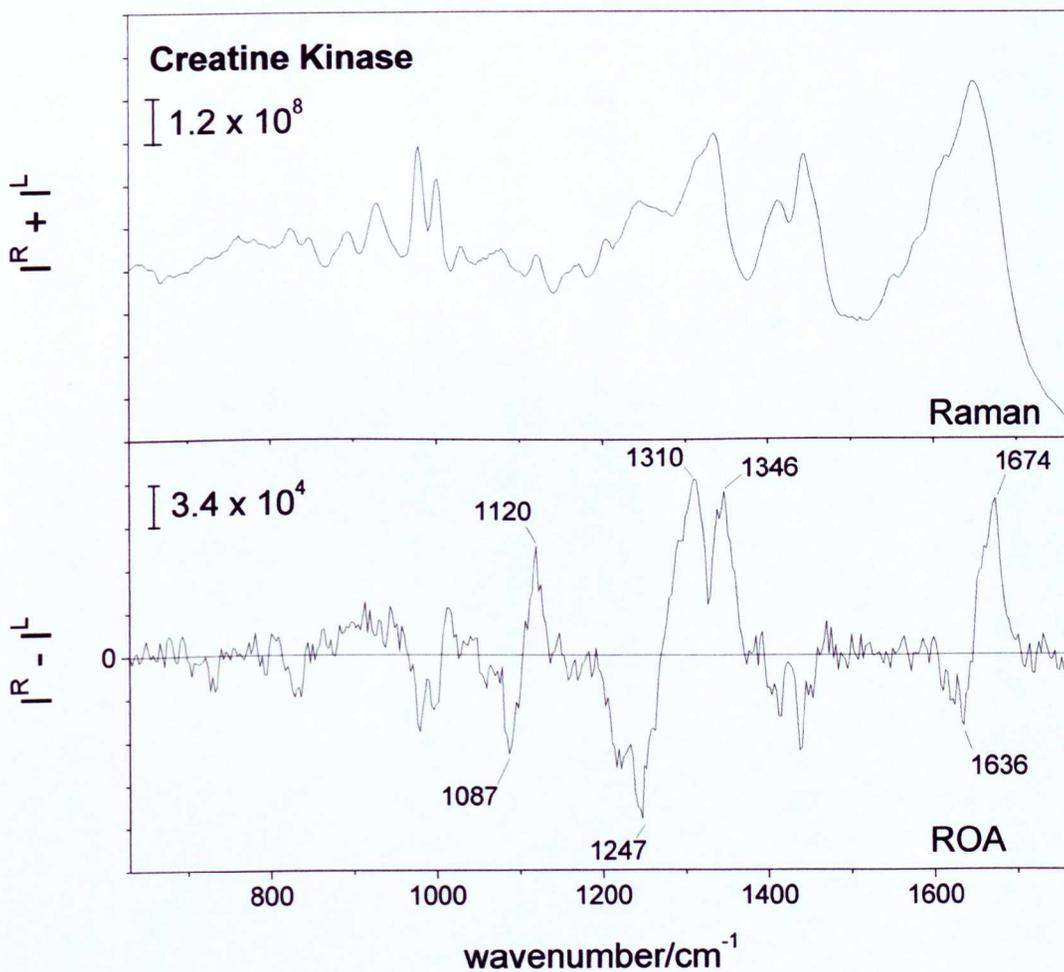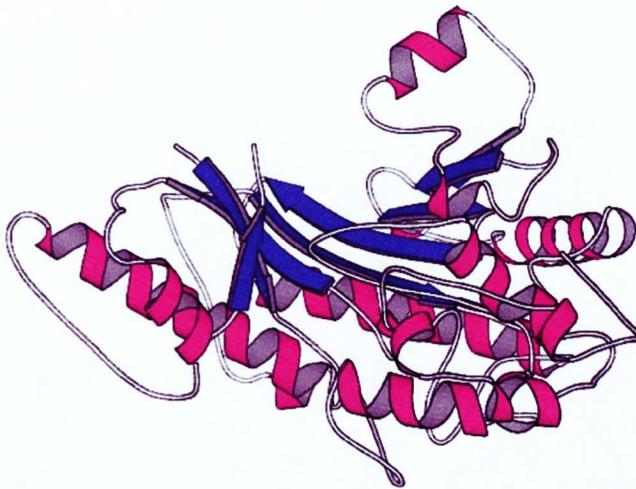
**Figure 3.8.** Molscript diagram and the backscattered Raman and ROA spectra of creatine kinase from rabbit muscle. Spectra were measured at ~100 mg/mL in 0.1M acetate buffer at ~pH 5.4.

in the amide III region. However, strong positive bands at ~1120 and ~1346 cm$^{-1}$ provide ample evidence of a large amount of α-helix. As one might expect, there is also significant evidence for the presence of β-sheet with a strong negative peak at ~1247 cm$^{-1}$, and also a slight shoulder at ~1292 cm$^{-1}$ on the left hand side of the positive ROA band peaking at ~1310 cm$^{-1}$. Also, the couplet in the amide I region with a strong sharp positive peak at ~1674 cm$^{-1}$ (broadening to the lower wavenumber side) suggests significant β-sheet as well as α-helical content This evidence from the ROA spectrum alone is enough to suggest that there is at least ~10% β-strand/sheet present in this protein, while the bulk of the secondary structure is α-helix.

## 3.4 Conclusion

The spectra discussed in this chapter highlight the ability for ROA to distinguish between secondary structure types where there is a definite abundance of one type over another, in this case α-helix over β-strand/sheet. The usefulness of ROA as a probe of protein structure and its future application into quantitative structure determination relies upon being able to make such distinctions. In addition, differences between fold characteristics of proteins of the same general structural class (i.e. mainly alpha) are more subtle than the differences between folds of proteins of other fold classes. The spectra shown in this chapter provide a good example of how the ROA spectrum of proteins with the same fold are recognisable and yet subtly different.

As will be shown in subsequent chapters, there are other examples of 'fold families' that also show a great deal of similarity on visual inspection as well as a range of proteins with the same structural elements as some of the proteins mentioned in this chapter but in different relative proportions (in particular the alpha beta proteins of Chapter 5). The aim of the analysis of these spectra is to illustrate that as well as providing a general overview of the structure of the protein in question, it is also feasible to elucidate more detailed structural information. From these spectra, it is possible to distinguish a variety of bands and band patterns that reveal not only the type of structures present, but also their relative quantities.

# Chapter Four
# Beta Proteins

## 4.1  Introduction

Of the many proteins with regular folds studied as part of this project, the single largest group of proteins with similar fold characteristics belong to the beta class. This large group, for the purposes of this discussion, encompasses all proteins whose major structural component (as described by their crystal or NMR structures) is β-strand. In the case of the samples discussed in this chapter, this means that each protein has between ~25-50% β-strand content. However, within the beta class of proteins, there are two distinct subgroups; mainly beta proteins, which as well as their beta content may also contain up to ~10% α-helix; all beta proteins, which contain very little or no α-helix. Despite this distinction, it is common to find that the architecture of mainly beta and all beta proteins are very similar, a fact that is clearly reflected in the similarity of the overall appearance of the ROA spectra of these proteins.

The major purpose of studying a number of mainly or all beta proteins is to attempt to solidify a number of important ROA band assignments that pertain to the variety of beta structures in native proteins. The study of the model α-helical structures (in the form of homopolypeptides) has been extremely useful in the assignment of various types of helical signatures in ROA spectra. Similarly, disordered polypeptides have proved useful in the assignment of loop and turn structure. However, the lack of suitable model beta structures (mainly due to gel formation) has presented a problem when attempting to assign meaning to observed ROA bands. Therefore, a survey of a group of mainly beta proteins, on the basis of a pre-knowledge of their structures from X-ray crystal structures or solution NMR structures, is a logical step in addressing this problem.

This chapter is presented in two sections. Firstly, the spectra of the all beta proteins of concanavalin A (from Jack bean), hen avidin, human immunoglobulin G and

human serum amyloid P component will be discussed. Secondly, the mainly beta proteins of α-chymotrypsin, β-lactoglobulin, trypsin and trypsinogen (all bovine) and human ubiquitin are presented.

## 4.2 Experimental

Samples of α-chymotrypsin, trypsinogen (bovine) and human immunoglobulin were supplied by Sigma-Aldrich Company Ltd. Concanavalin A, hen avidin and human ubiquitin were supplied by Fluka (Sigma-Aldrich Company Ltd.). Bovine trypsin was supplied by Calbiochem Ltd. Bovine β-lactoglobulin was supplied by ICN Ltd. Human serum amyloid P component was prepared and supplied by Dr. G. Tennent of the Royal Free and University College Medical School at University College London.

## 4.3 Results and discussion

### 4.3.1 All beta proteins

In this section, an independent discussion of each protein ROA spectrum is presented, and a comparison of all the spectra is undertaken and discussed separately.

## Concanavalin A

Concanavalin A is a member of the lectin family and is a carbohydrate binding protein consisting of four identical monomers, each with a molecular weight of ~30 kDa (Edmundson et al., 1971; Wang et al., 1971). Concanavalin A forms a jelly-roll β-barrel consisting of eight β-strands in an antiparallel β-sheet arrangement (Branden & Tooze, 1991; Weisgerber & Helliwell, 1993). As described on page 22, the jelly roll structure is comprised of two β-sheets with little hydrogen bonding between strands from different sheets, forming two Greek key motifs. The crystal structure of concanavalin A from Jack bean is available (PDB code 2cna) (Reeke et al., 1975) and is illustrated as a Molscript
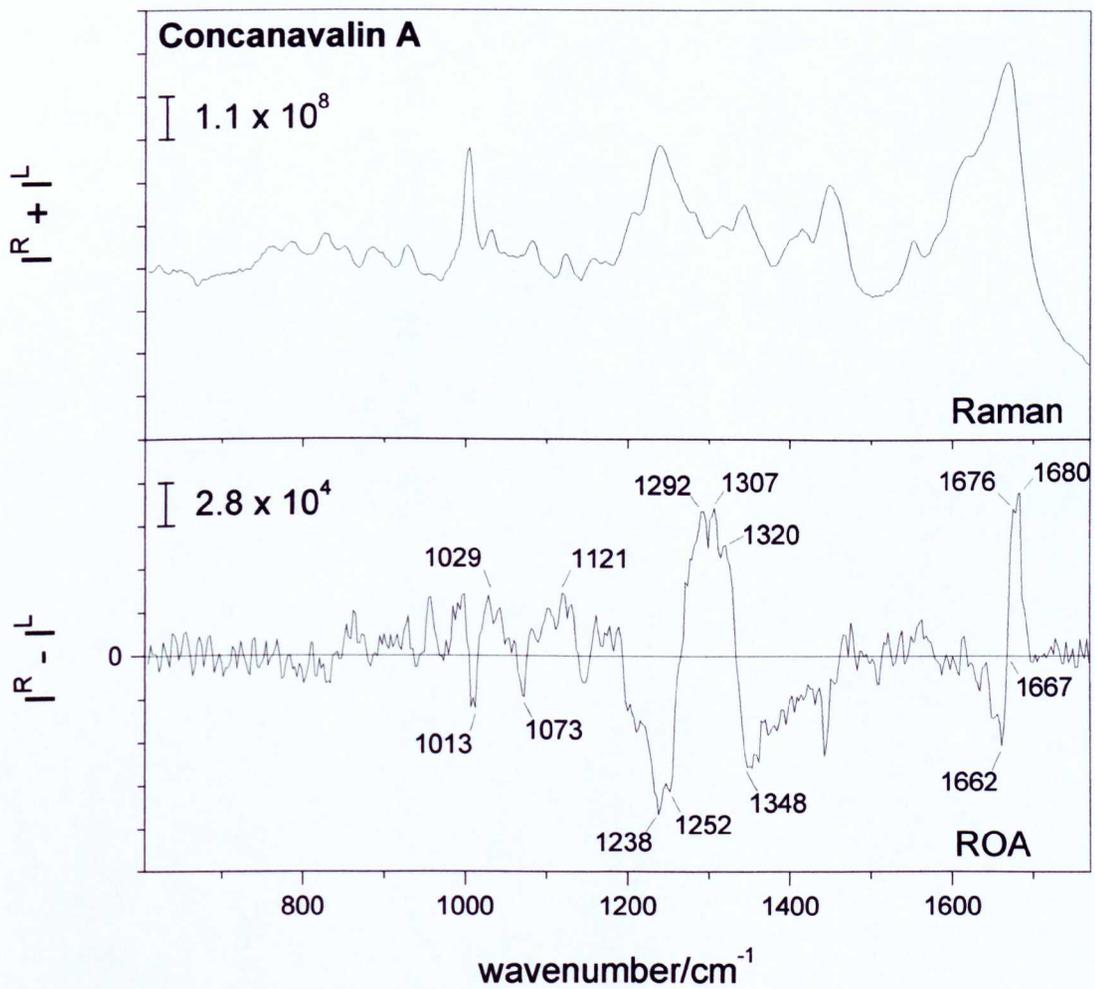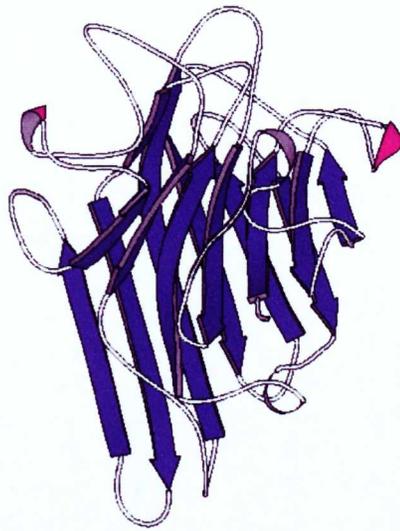
**Figure 4.1.** Molscript diagram and backscattered Raman and ROA spectra of concanavalin A (from Jack bean) in 0.1M acetate buffer at pH 5.4, ~67 mg/mL.

diagram in Figure 4.1. The monomer unit (beta domain) is 237 residues in length and contains less than 2% $\alpha$-helix and also less than 2% $3_{10}$-helix.

The backscattered Raman and ROA spectra of concanavalin A (from Jack bean) are also shown in Figure 4.1. The amide I region immediately suggests a protein with high $\beta$-sheet content, with a strong, narrow couplet with a negative peak at ~1662 cm$^{-1}$ and positive peak at ~1680 cm$^{-1}$. The amide III region is typical of a $\beta$-barrel protein, with a tall, slender positive region flanked by negative structure to both high and low wavenumber. There is clear negative structure incorporating peaks at ~1220 and ~1245 cm$^{-1}$ to reflect the high $\beta$-strand content of the protein, and negative structure peaking at ~1348 cm$^{-1}$ from $\beta$-hairpins. Also in the amide III region, there is a clear lack of any characteristic ROA signatures of $\alpha$-helix. Instead, there is positive structure peaking at ~1292 cm$^{-1}$, ~1307 cm$^{-1}$ and ~1320 cm$^{-1}$, where the ~1292 cm$^{-1}$ band is possibly reflecting the residues in longer loops with $\phi,\varphi$ angles in the general $\alpha$ region of the Ramachandran plot. The peak at ~1307 cm$^{-1}$ is thought be generated by the presence of $\beta$-turns (Wilson, 1996). The ~1320 cm$^{-1}$ band (as will be described in more detail in subsequent chapters) is likely indicative of PPII structure within extended loops. As the X-ray crystal structure illustrates, the jelly roll $\beta$-barrel has a large number of connecting loops between individual $\beta$-strands.

**Hen avidin**

Avidin is a biotin-binding protein and is a minor constituent of hen egg white. Taking its name from the avidity to which it binds with biotin, it binds so strongly that it is extremely difficult to separate the two molecules (Torreggiani et al., 1998). Its biological function is unknown, but it binds one molecule of biotin per subunit. The protein is a homotetramer, with each monomer subunit having a molecular weight of ~16kDa. The X-ray crystal structure of hen avidin is available (PDB code 1avd) (Pugliese et al., 1984) and reveals an eight stranded anti-parallel beta-barrel enclosing an internal ligand binding site. The three dimensional structure is represented as a Molscript diagram in Figure 4.2. Again, the X-ray structure reveals the highly β character of this protein, with almost 50% β-strand, 0% α-helix and ~ 2.5% $3_{10}$-helix.

The backscattered Raman and ROA spectra of hen avidin in aqueous solution are shown in Figure 4.2. Again, the most prominent features of the ROA spectrum reveal a protein with highly β character. The most notable indicators are the exceptionally strong positive amide I peak at ~1675 cm$^{-1}$, forming a couplet with a negative peak to the lower wavenumber side, reflecting the β-sheet content. The amide III region is also highly suggestive of β character, with clear negative peaks at ~1217 and 1239 cm$^{-1}$ from β-strand, and negative structure peaking at ~1360 cm$^{-1}$ indicating the presence of a number of β-hairpins.

Once again, the positive structure seen in the amide III region has a peak at ~1285 cm$^{-1}$, which is ~10-15 cm$^{-1}$ lower than those seen commonly in helical proteins. As in the previous case, it is thought that this peak may be attributable to residues in longer loops with φ,φ angles in the general α region of the Ramachandran plot, but may well be coupled to β-type structure frequently observed from ~1245 cm$^{-1}$ and upward. The other peak in the amide III region is centred at ~1310 cm$^{-1}$ and is arguably too low to be assigned to PPII structure. However, it is not a clearly defined peak in so much as it is a small peak on top of a broader, underlying structure which may well have a significant boost from PPII structure from between ~1315-1320 cm$^{-1}$.
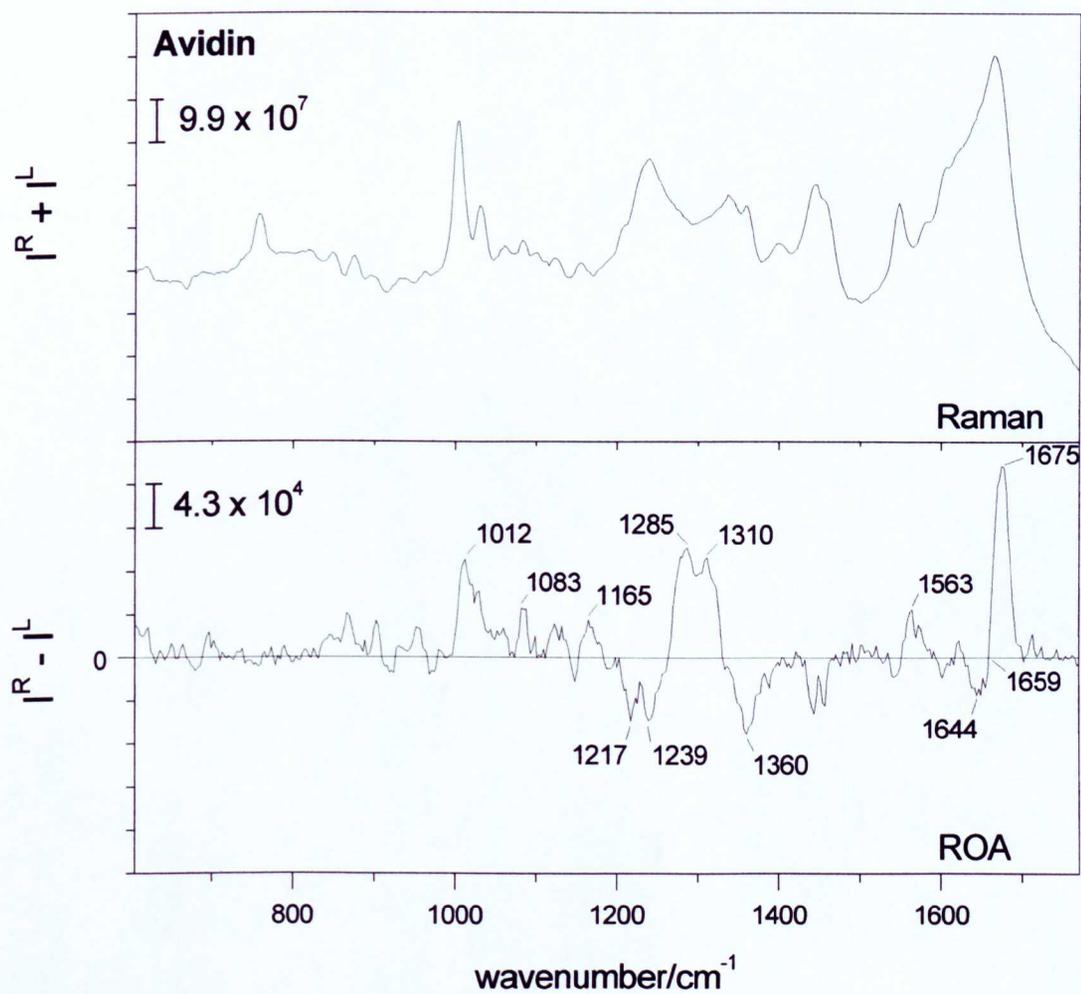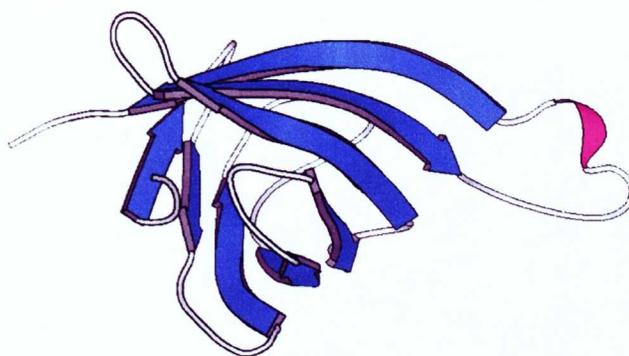
**Figure 4.2.** Molscript diagram and backscattered Raman and ROA spectra of hen avidin in 0.1M acetate buffer at pH 5.4, ~67 mg/mL.

## Human immunoglobulin G

Human immunoglobulin G (IgG) is a glycoprotein with a molecular mass of ~150 kDa, and contains ~5% carbohydrate content. The basic structure of the human immunoglobulin molecule is a tetramer of two light chains and two heavy chains linked by disulphide bonds, and is depicted in Figure 4.3A. Each chain is made up of
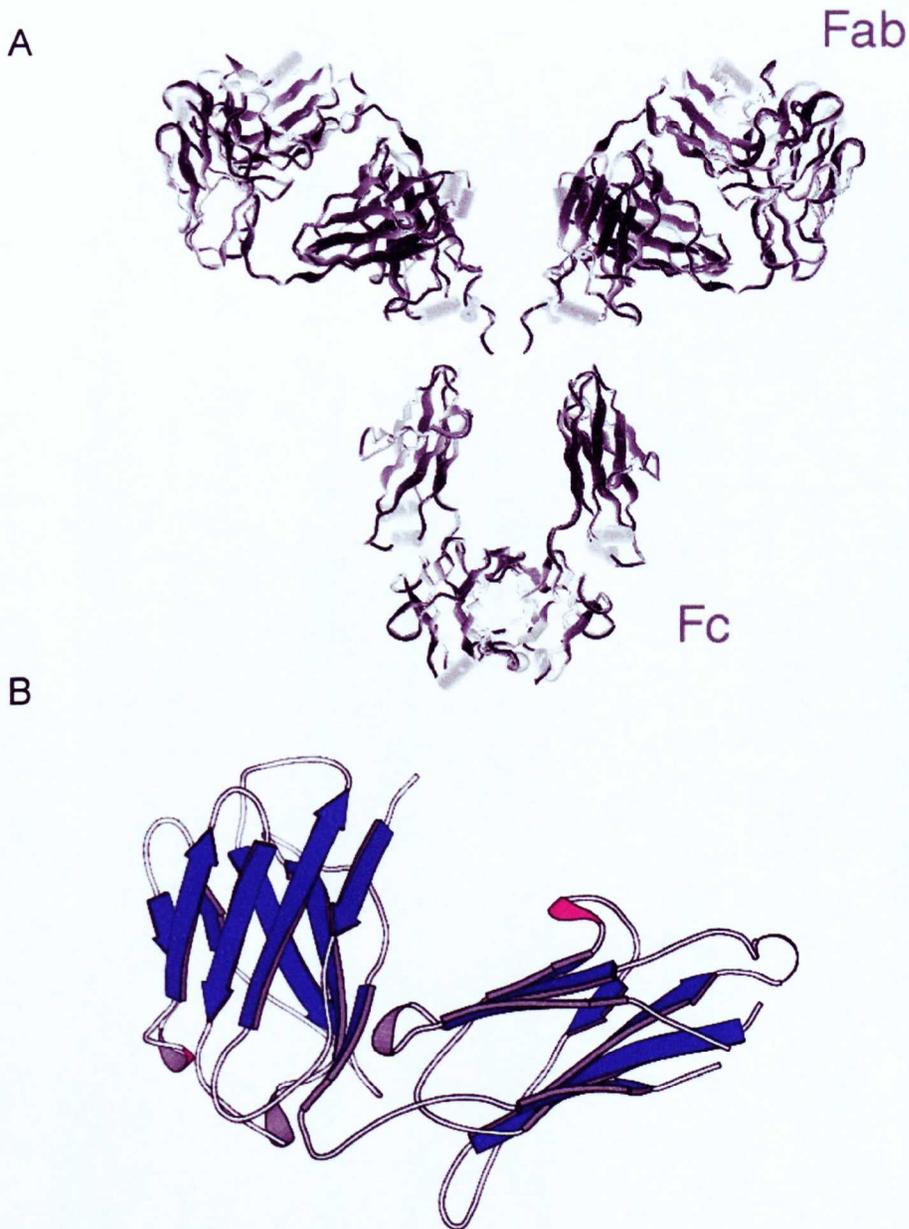


**Figure 4.3 A** –Human immunoglobulin G tetramer, showing the $F_{ab}$ and $F_c$ regions. **B** –Molscript diagram of the X-ray crystal structure (PDB code 7fab) of human immunoglobulin G (lambda light chain (H)).

one (light chain) or two (heavy chain) subunits as shown in Figure 4.3B. The crystal structure of the human IgG subunit has been solved (Saul et al., 1991) and represents one of the most common superfold types as described by Orengo et al. (1997) in the CATH classification system. Comprising a β-barrel structure with a Greek-key motif and consisting of antiparallel β-sheet, its crystal structure predicts a β-strand content of ~49%, and less than 5% α-helix.

The backscattered Raman and ROA spectra of human immunoglobulin G in aqueous solution are shown in Figure 4.4. As seen in previous spectra in this section,
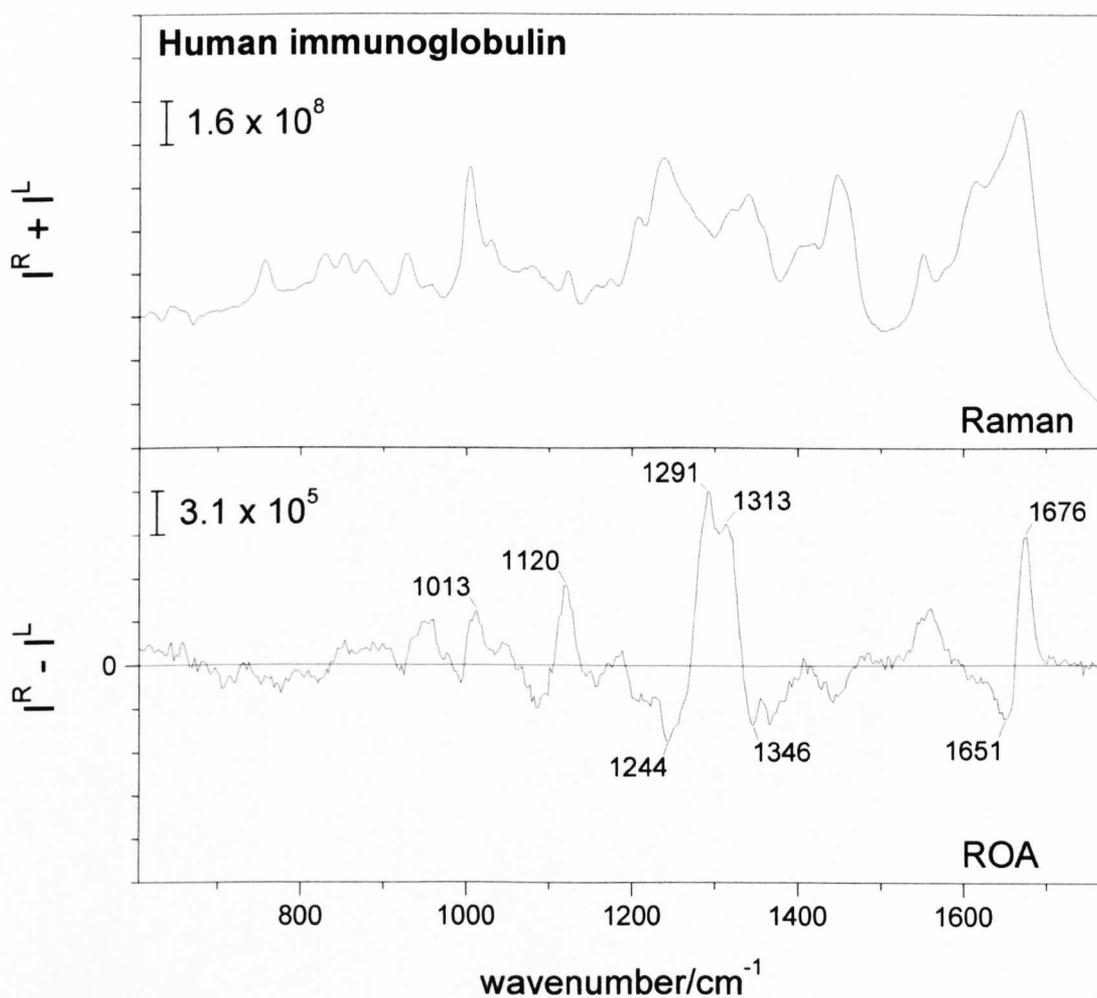


**Figure 4.4.** The backscattered Raman and ROA spectra of human immunoglobulin G in 0.1M acetate buffer at pH 5.4, 100mg/mL.

the ROA spectrum is highly suggestive of β structure, with the amide I couplet showing a strong, sharp positive peak at ~1676 cm$^{-1}$, and a negative peak relatively high at ~1651 cm$^{-1}$. Again, negative structure at ~1244 cm$^{-1}$ and ~1346 cm$^{-1}$ signify the presence of β - strand and hairpins respectively. There is a clear positive peak at ~1291 cm$^{-1}$ in the amide III region, again possibly due to helical residues in loops and turns, but conventional Raman studies suggest that types I and III β-turns may also contribute significantly in this region (Krimm & Bandekar, 1980/1986). The positive ROA band at ~1313 cm$^{-1}$ is similar to bands seen in the previous examples in so much as it is not in itself a clear peak, and the intensity at ~1320 cm$^{-1}$ maybe counter-acted by the significant contribution from β-hairpins that give rise to the negative structure at ~1346 cm$^{-1}$. As a consequence, it may be valid to assign this peak to the presence of PPII structure.

**Human serum amyloid P component**

Serum amyloid P component (SAP) is a precursor of amyloid P component which is associated with lysozyme amyloidosis and a wide range of other diseases which involve large quantities of amyloid deposits (Pepys et al., 1997). As such, the study of the structure of this protein, as well as the many others that are implicated in diseases where amyloidosis occurs is of great current interest. SAP is a glycoprotein that binds in a calcium-dependent fashion to a variety of ligands including other proteins and DNA (Hohenester et al., 1997). It is a glycoprotein composed of a pair of noncovalently bound pentameric discs with a subunit size of 23-25 kDa. Each subunit consists of 204 amino acid residues, a single disulfide bridge, and a carbohydrate moiety (Castano et al., 1985).

The crystal structure of the SAP monomer is available (PDB code 1lgn) (Hohenester et al., 1997) and is illustrated in Figure 4.5. It describes a highly β structure, with ~44% β-strand, ~4% α-helix and ~1.5% $3_{10}$-helix. The diagram in Figure 4.5 also illustrates the fact that there is also a significant amount of long loops connecting the β-strands, which would be expected to be evident in the ROA spectrum of the protein.
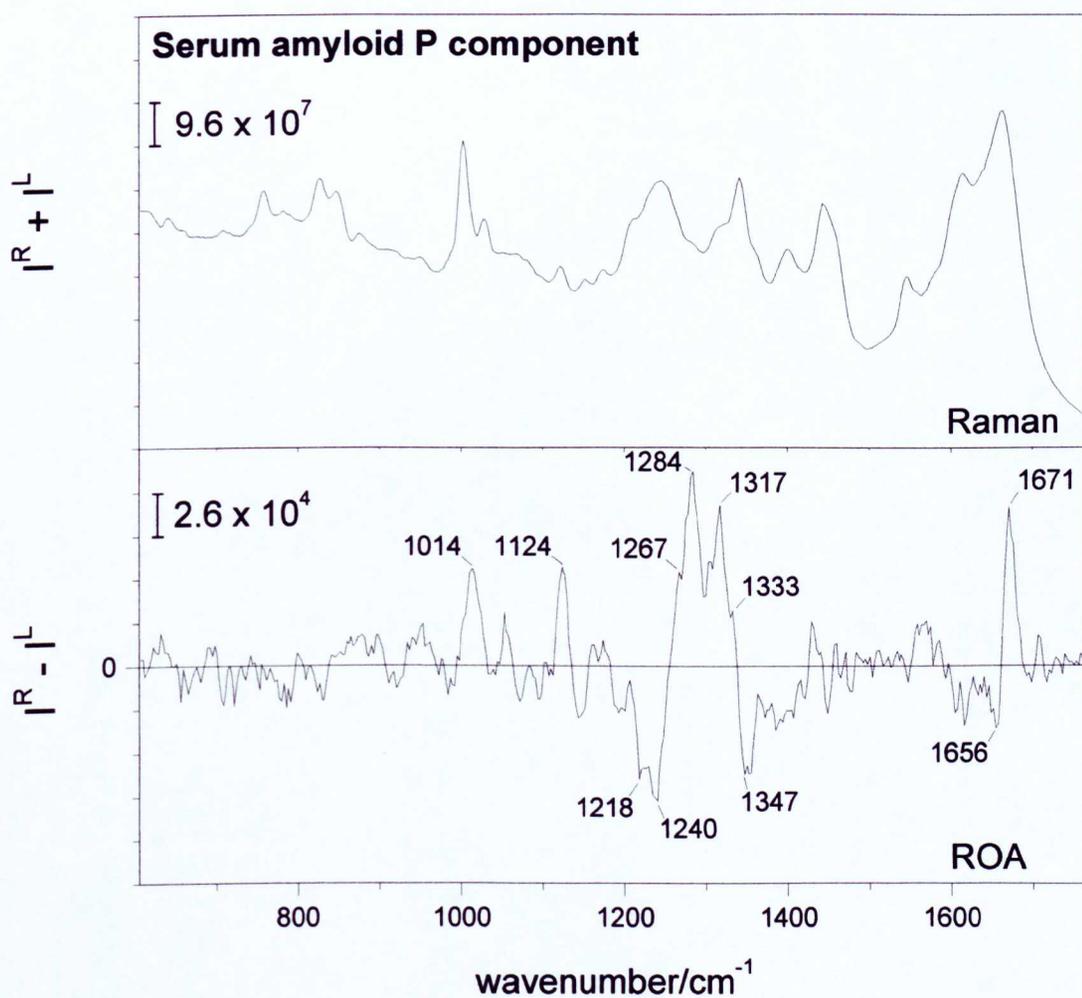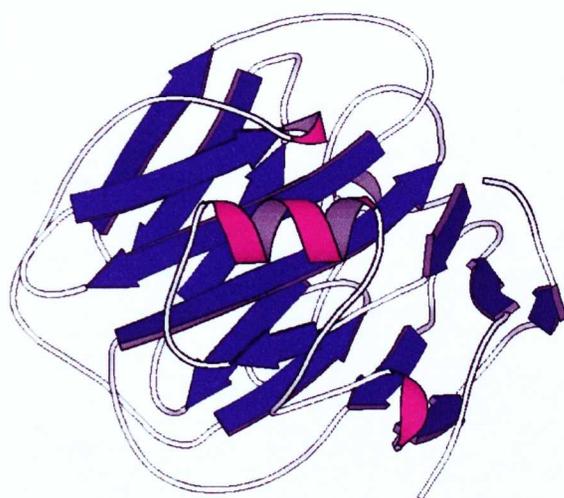
**Figure 4.5.** Molscript diagram and the backscattered Raman and ROA spectra of human serum amyloid P component in 0.1M acetate buffer at pH 5.4., ~100 mg/mL.

The backscattered Raman and ROA spectra of human serum amyloid P component are shown in Figure 4.5. In common with the other three samples discussed in this section, the ROA spectrum reflects the highly β nature of the protein. Specifically, the negative-positive couplet in the amide I region once again is strong and sharp, and shifted to higher wavenumber, with a positive peak at ~1671 cm$^{-1}$ and a negative peak at ~1656 cm$^{-1}$. The amide III region also shows peaks that have been assigned to β-strand between ~1218 and ~1240 cm$^{-1}$, and β-hairpins at ~1347 cm$^{-1}$. However, the positive structure in the amide III region of this protein is particularly well defined and shows a very strong band at ~1284 cm$^{-1}$, similar to that seen in avidin but much more intense. The assignment of this band is still unclear, but the length of the loops in SAP suggests that there are fewer turns in proportion to the length of the polypeptide chain, and that it is conceivable that the band at ~1284 cm$^{-1}$ is due to the presence of residues within these loops that have $\phi,\phi$ angles corresponding to the same region within the Ramachandran plot. The other most prominent feature of this spectrum is the strong, sharp positive ROA band at ~1317 cm$^{-1}$. As anticipated by the crystal structure, and from analysis of the Ramachandran plot, the extended loop structure is likely to contain sections of PPII helix (see Chapter 6 for a more detailed discussion of PPII helix), which is believed to generate positive ROA intensity at ~1320 cm$^{-1}$ (Syme et al., 2002).

### 4.3.2 Summary of all beta proteins

The similarity of these four spectra shows that ROA is not only capable of distinguishing individual structural elements within proteins, but constructs a picture of the molecule as a whole. Therefore, it can be said that the pattern of bands in the ROA spectrum of a protein reflects the tertiary fold. (Barron et al., 2000) The visual similarity of these spectra is emphasized in Figure 4.6, which shows all four spectra together, and the most prominent common bands marked. The proteins included in this section are but the most recent additions to a growing set of data that, with their common spectral features and experimental reproducibility, are useful in solidifying the assignment of physical significance to ROA spectral bands.
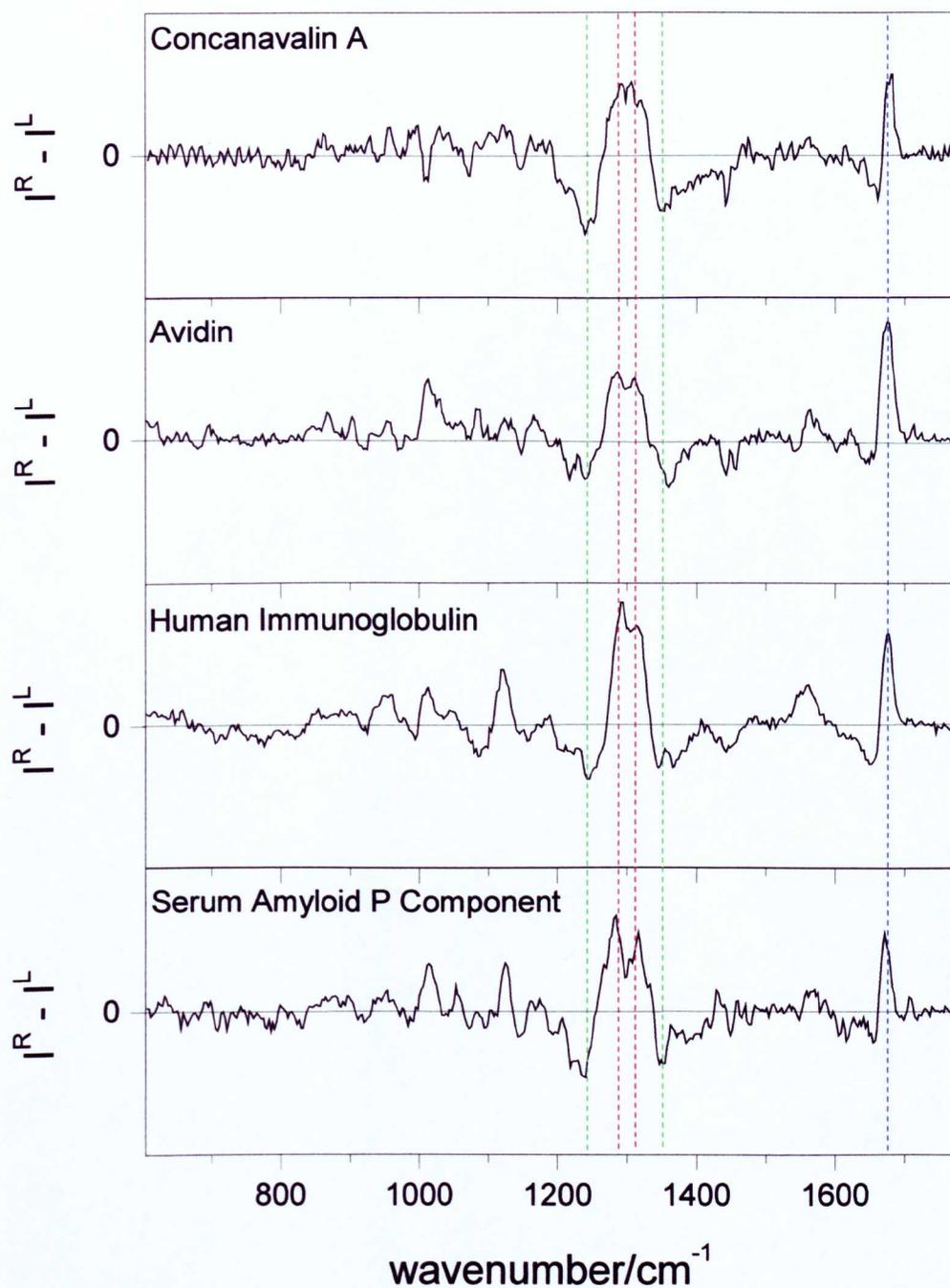
**Figure 4.6.** The backscattered ROA spectra of four all beta proteins. Top to bottom – Concanavalin A, hen avidin, human immunoglobulin and human serum amyloid P component. Common bands are highlighted for clarity. The positive amide III peaks at ~1290 and ~1310 cm$^{-1}$ are shown in red, the negative amide III peaks at ~1240 and ~1345 cm$^{-1}$ are shown in green and the positive amide I peak at ~1676 cm$^{-1}$ is shown in blue.

## 4.3.3  Mainly beta proteins

The proteins in this section generally have between 20-40% β-strand, with up to 10% α-helix as well. Firstly, three members of the trypsin family are discussed. The ROA spectra of β-lactoglobulin above and below the Tanford transition (that occurs around pH 7.5) are then presented, as well as human ubiquitin. All the proteins in this section, with the exception of human ubiquitin, are β-barrel proteins similar to those shown in the previous section.

**Bovine α-chymotrypsin**

Bovine α-chymotrypsin is a serine protease and a member of the trypsin family. Members of the chymotrypsin family may occasionally function intracellularly, for example in the intracellular digestion of bacteria, but most function extracellularly, for example in food digestion (Rawlings et al., 1994). α-chymotrypsin has a molecular weight of ~25 kDa and is 237 residues long. The crystal structure is available (PDB code 2cha, Birktoft & Blow, 1972) and is illustrated as a Molscript diagram in Figure 4.7. The crystal structure suggests ~32% β-strand and ~7% α-helix, in the form of a closed β-barrel with anti-parallel β-strands forming a Greek-key motif. The protein consists of two domains of the same fold, as illustrated in the Molscript diagram.

The backscattered Raman and ROA spectra of bovine α-chymotrypsin are shown in Figure 4.7. The overall appearance of the spectrum is, as one would expect, similar to that of β-barrel proteins discussed in the previous section, although the negative structure flanking the central positive structure in the amide III region is not so well defined and there is less intensity. The amide I region has a slightly less definitive appearance as the all β proteins, with a slightly broader positive peak, peaking at between ~1665 and ~1672 cm$^{-1}$. This suggests that the higher percentage of α-helix present is shifting the amide I couplet to a slightly lower wavenumber region. There is also evidence of some α-helix due to the small positive peak at ~1300 cm$^{-1}$, and in the structured part of the skeletal
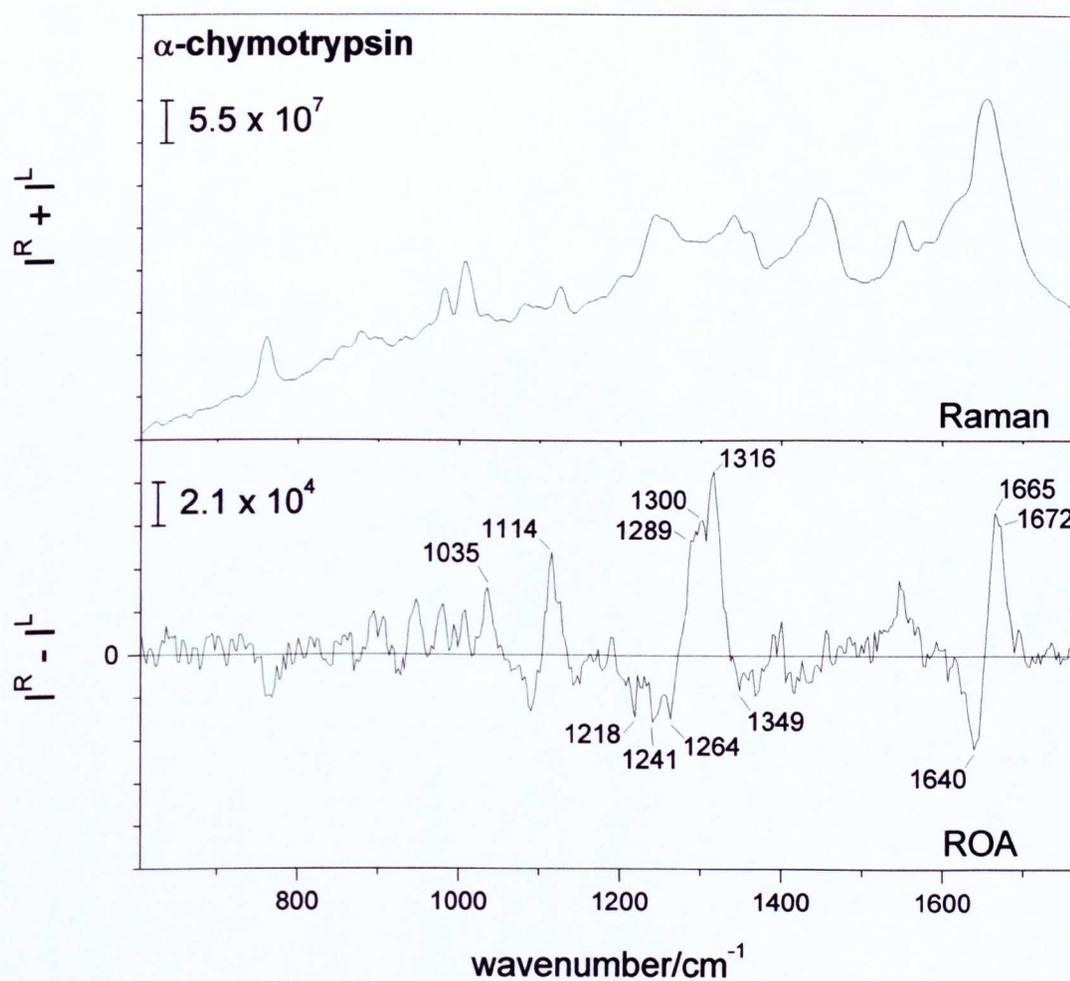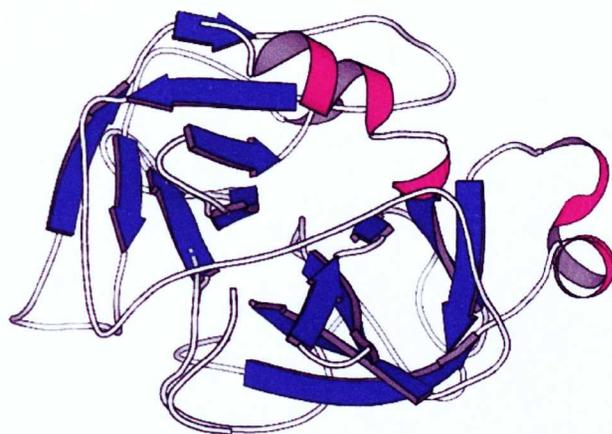
**Figure 4.7.** Molscript diagram and the backscattered Raman and ROA spectra of bovine α-chymotrypsin in 0.1M acetate buffer with ~8mM KI solution at pH 5.4, 100mg/mL.

backbone skeletal stretch region at ~950 cm$^{-1}$. The strongest peak in this spectrum is at ~1316 cm$^{-1}$, similar to human SAP and immunoglobulin G. The appearance of the whole spectrum is, in general, most similar to that of human immunoglobulin G, with a common peak at ~1114 cm$^{-1}$. This feature, however, is conspicuous in many proteins, including many α+β proteins and is as yet unassigned.

**Bovine trypsin**

Trypsin from bovine pancreas is also a serine protease and another member of the trypsin-like family proteins with similar fold characterstics to α-chymotrypsin and trypsinogen. It also has a molecular weight of ~25 kDa and is 224 amino acid residues in length. The crystal structure (PDB code 1try, Rypniewski et al., 1994) of trypsin from mold reveals a closed β-barrel structure with two domains of the same fold type, with β-strands connecting in a Greek-key motif. Trypsin is known to contain ~33% β-strand, and ~10% α-helix, and its backscattered Raman and ROA spectra are shown in Figure 4.8. Despite the similarity in the fold characteristics between trypsin and α-chymotrypsin, the ROA spectra are quite distinctly different. Trypsin has a broader amide I couplet, yet both the negative and positive peaks are in roughly the same positions as in the previous example. The broadening of the amide I band is probably due to the presence of increasing amounts of α-helix, and also influenced by the amount of random coil present in the structure. The main β signatures in the ROA spectrum of trypsin are the negative bands at ~1220 cm$^{-1}$, currently assigned to β-strand and a weak peak at ~1243 cm$^{-1}$.

The main feature in the amide III region is the positive intensity that peaks at ~1308 cm$^{-1}$, with a shoulder at ~1302 cm$^{-1}$, which is reflecting the α-helix content of the protein. This band in particular suggests that the α-helices are in a hydrophobic environment. The peak at ~1308 cm$^{-1}$ is similar to that seen in concanavalin A (see Figure 4.1) and is assigned to β-turns (Wilson, 1996; Barron et al., 2000).
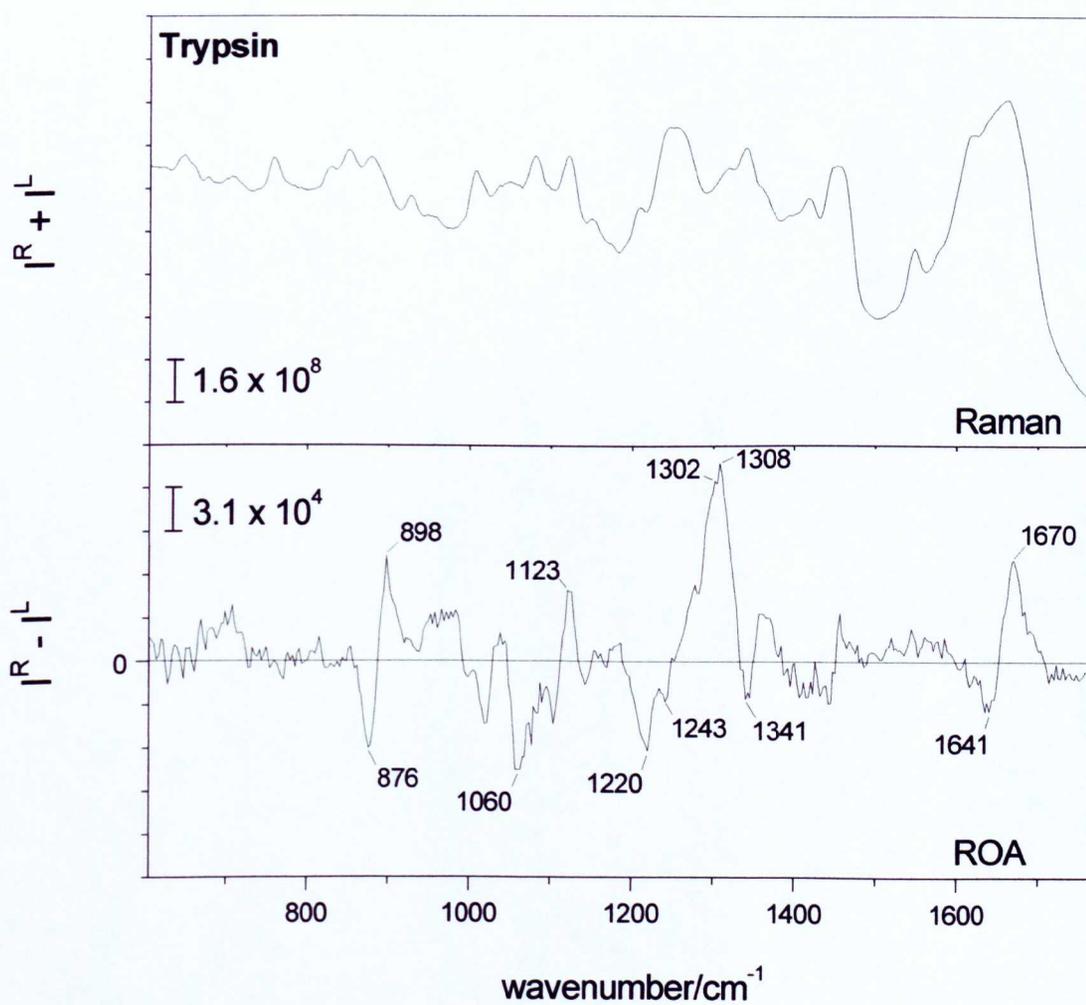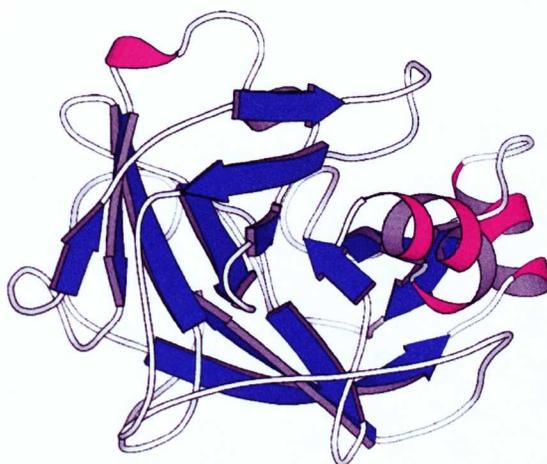
**Figure 4.8.** Molscript diagram and the backscattered Raman and ROA spectra of trypsin from bovine pancreas in 0.1M acetate buffer with ~8mM KI solution at pH 5.4, ~100 mg/mL.

# Bovine trypsinogen

Bovine trypsinogen is also a serine protease and the third member of the trypsin family in the current data set to be analyzed. Trypsinogen is the inactive precursor of trypsin with molecular weight ~25 kDa, consisting of 222 amino acid residues. Trypsinogen contains three disordered loop sections 7-12 residues long that are ordered in trypsin. The rest of the molecules (~85%) are identical to each other. The crystal structure (PDB code 1tgn, Kossiakoff et al., 1977) is available and once again reveals a closed β-barrel structure with strands connecting in a Greek-key motif, as shown in Figure 4.9. as a Molscript diagram.

The backscattered Raman and ROA spectra of bovine trypsinogen are also shown in Figure 4.9. There is a noticeable similarity to the previous β-barrel proteins shown in this chapter, with a number of very similar band assignments. However, the detailed structure of the ROA spectrum tends to more closely resemble that of the mainly β proteins, as opposed to the all β proteins in the previous section. Notably, the negative structure flanking the central positive structure in the amide III region to low and high wavenumber are not as pronounced, as a consequence of its lower overall β-strand content (~33%). The peaks of greatest interest in this spectrum once again originate from the dominant β-strand content, with a sharp peak at ~1217 cm$^{-1}$ and ~1253 cm$^{-1}$. The positive peaks in the amide III region at ~1299 cm$^{-1}$ and ~1309 cm$^{-1}$ represent the α-helical and turn content respectively. However, there is a suggestion that the positive intensity to the right of the band at ~1309 cm$^{-1}$ (which shows a shoulder at ~1330 cm$^{-1}$) is reflecting either a small amount of hydrated α-helix (which would generate positive ROA at ~1335-1345 cm$^{-1}$) which in turn could be counteracting the negative ROA contribution from β-turns and hairpins at ~1345-1360 cm$^{-1}$.

The ROA spectra of the trypsin-like serine proteases represent a unique insight into the different spectral characteristics of three evolutionarily linked proteins. Despite very similar secondary structure contents and tertiary fold characteristics, the overall
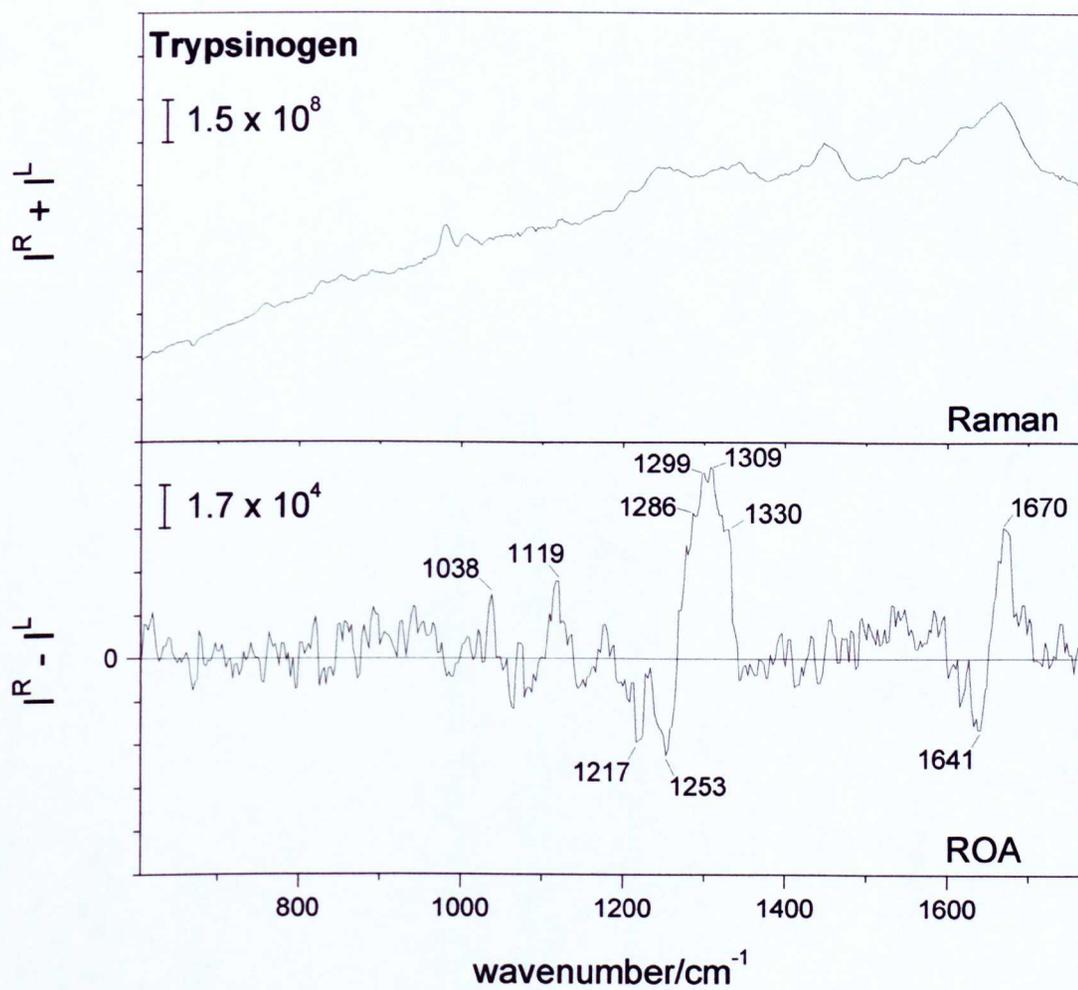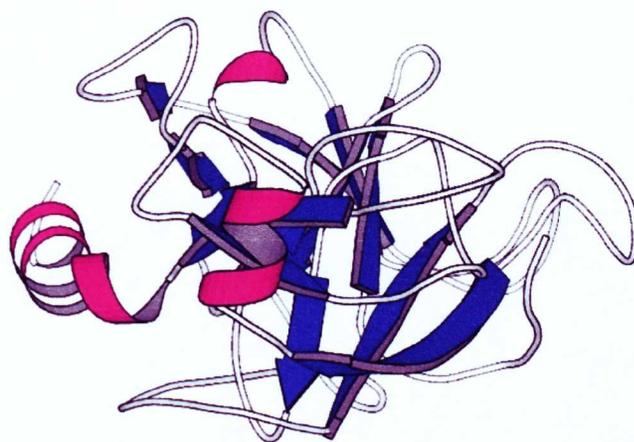
**Figure 4.9.** Molscript diagram and the backscattered Raman and ROA spectra of bovine pancreas trypsinogen in 0.1M acetate buffer at pH 5.4., ~100 mg/mL.

appearance of the ROA spectra differ to the point where each spectrum is quite distinct from the others, yet it is mainly the relative intensities of the bands and not the position of the bands that differs from sample to sample. The differences seen in these spectra are quite marked when they are considered together in this manner, but when taken as part of a more general survey (i.e. when analyzed in context of all the ROA spectra discussed in this thesis), the similarities far outweigh the differences. However, the task of quantifying these similarities and differences is still a fairly subjective one, and this problem is addressed in the final chapter of this thesis.

### β-lactoglobulin

The mammalian milk protein β-lactoglobulin is a lipid-binding protein and a member of the lipocalin family (Sawyer, 1987; Banaszak et al., 1994). Its function is still unclear but it binds with and is probably involved with the transport of retinol. β-lactoglobulin has been the subject of many physicochemical studies and, as well as oligomerization, it undergoes interesting conformational changes with changes in pH, temperature and concentration. For this reason, the study of this protein is of great importance to the subject of protein folding in general and has recently been the subject
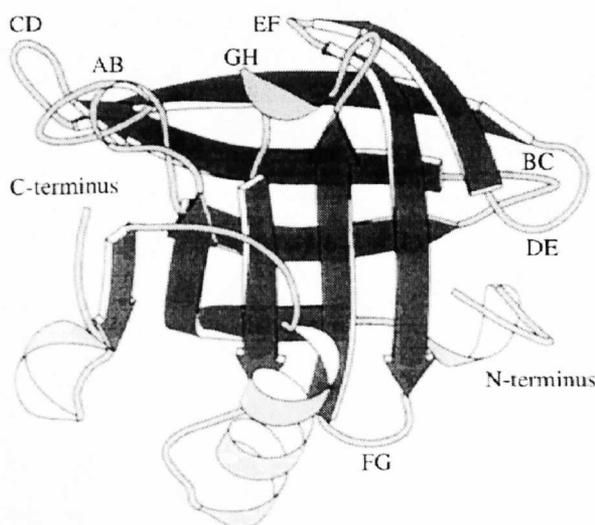


**Figure 4.10.** Diagram of one subunit of b-lactoglobulin lattice X. The data used to generate this diagram is originally from Brownlow et al. (1997) and is in the PDB (code 1beb)

of a detailed ROA study (Blanch et al., 1999). The structure of one subunit of the protein molecule is shown in Figure 4.10 and was reproduced from Blanch et al. (1999). The letters denote the strands being linked by the individual loops. The study of β-lactoglobulin has been central to the understanding of several key β-type assignments in the ROA spectra of proteins. The crystal structure of bovine β-lactoglobulin (PDB code 3blg) is available (Qin et al., 1998) and predicts an up-down β-barrel core (containing ~40% β-strand) with 10-12% α-helix. The protein exists as a dimer at neutral pH, and dissociates into monomers below ~pH 3. The monomer has a molecular weight of around ~18 kDa, and is composed of 162 amino acid residues and contains two disulfide bonds.

β-lactoglobulin undergoes significant conformational changes at ~pH 7.5. This is known as the Tanford transition (Tanford et al., 1959) and is associated with the change in conformation of loops, without affecting the basic β-barrel structure (see Blanch et al., 1999 for a review). For the purposes of this project, the backscattered Raman and ROA spectra of bovine β-lactoglobulin were collected above and below the Tanford transition at pH 8.5 and pH 6.5 respectively, and these spectra are shown in Figure 4.11. It is evident that the basic structures of both spectra are very similar. They both show the characteristic β-barrel pattern (as seen in the previous spectra in this chapter) of a negative-negative-positive-negative band pattern in the amide III region, with peaks at ~ 1220, ~1245, ~1317 and ~1360 cm$^{-1}$ respectively. Also, the two spectra have very similar amide I couplets with a strong sharp positive peak above ~1670 cm$^{-1}$ and a negative peak at around ~1650 cm$^{-1}$. The similarity of this feature alone is a good indication of the unchanged conformational state of the β-barrel core of the subunit. In addition, there is great similarity in the skeletal backbone region (~1100 cm$^{-1}$ and below) with a pattern consistent with that observed in proteins with some helical content, specifically showing peaks at ~950 cm$^{-1}$.

However, it is the difference between the two spectra that is most interesting. The sharp positive ROA band of medium intensity at ~1268 cm$^{-1}$ present in the spectrum of β-
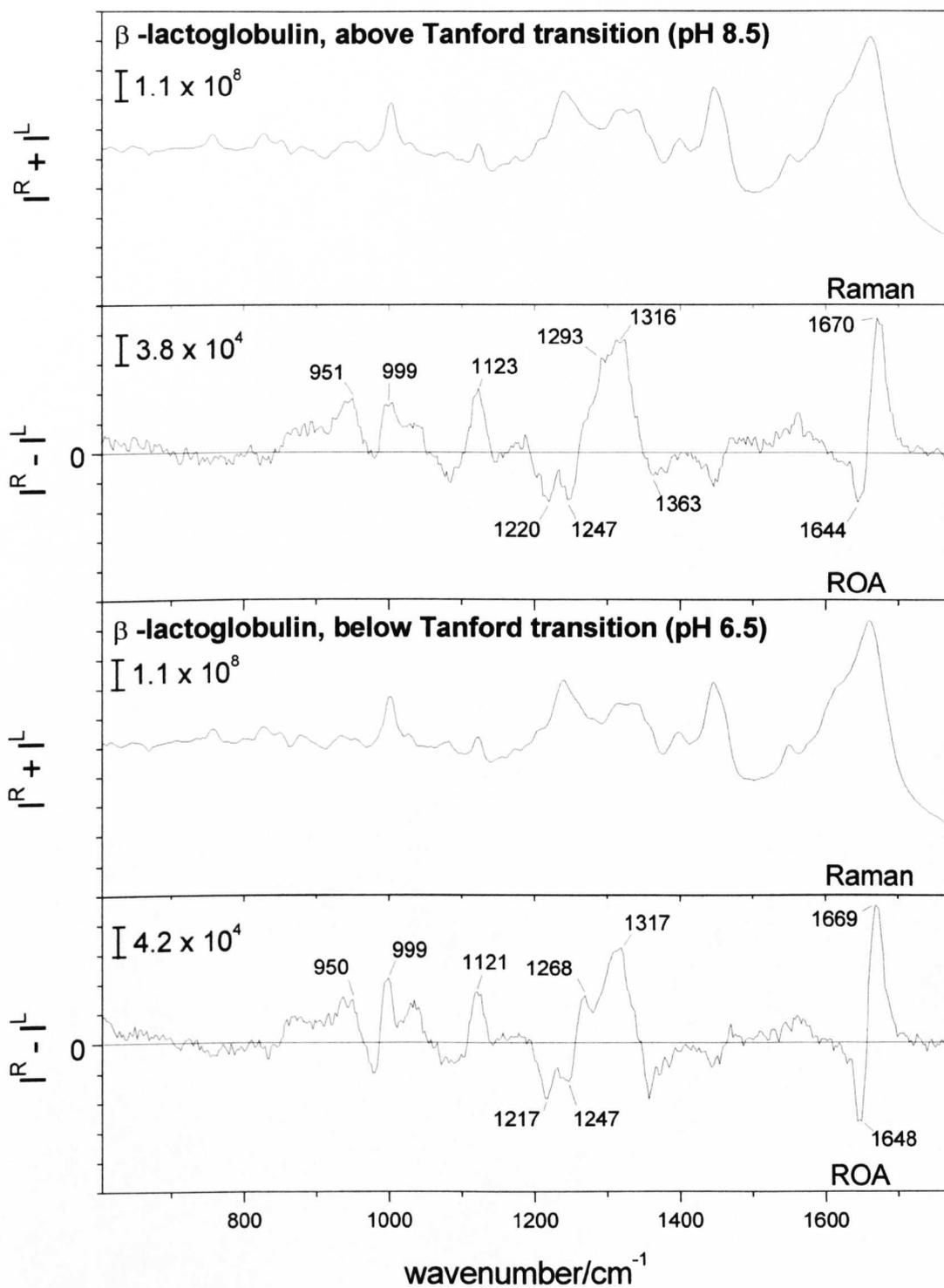
**Figure 4.11**. The backscattered Raman and ROA spectra of β-lactoglobulin above (top pair) and below (bottom pair) the Tanford transition.

lactoglobulin below the Tanford transition (at pH 6.5) is shifted to ~1293 cm$^{-1}$ on moving to higher pH (at pH 8.5, above the Tanford transition). This is discussed in detail in the paper by Blanch et al. (1999) and is beyond the scope of this current chapter. For the purposes of this discussion however, these new spectra represent the best quality data collected for this protein to date, and confirm the experimental observations made previously by Blanch et. al.

## Human ubiquitin

Ubiquitin is a chromosomal protein found in all eukaryotic cells whose sequence is extremely well conserved from protozoan to vertebrates (Wiborg et al., 1985). Ubiquitin is a globular protein with a molecular mass of ~8.5 kDa composed of 76 amino acids. It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response and ribosome biogenesis (Wiborg et al., 1985). Its crystal structure is available (PDB code 1ubi, Vijay-Kumar et al., 1987) and the ubiquitin-like β-grasp structure represents another of the nine major superfold classes as described by Orengo et al. (1997). The crystal structure of human ubiquitin is shown as a Molscript diagram in Figure 4.12. Ubiquitin is classed by the SCOP classification system (Murzin et al., 1995) as an α+β, as it contains a significant proportion of both types of structure. However, it is predominantly β in structure with ~32% β-strand (similar to the other proteins discussed in this section), but also has ~15% α-helix. Ubiquitin forms a β-grasp structure that is composed of a four-stranded curved β-sheet that wraps around a large α-helix.

The backscattered Raman and ROA spectra of human ubiquitin are shown in Figure 4.12. There is an immediate visual similarity to the β-barrel proteins but with one or two significant differences. Curiously for a protein with such a significant proportion of α-helix, the amide I couplet is surprizingly similar to that seen in even the all β
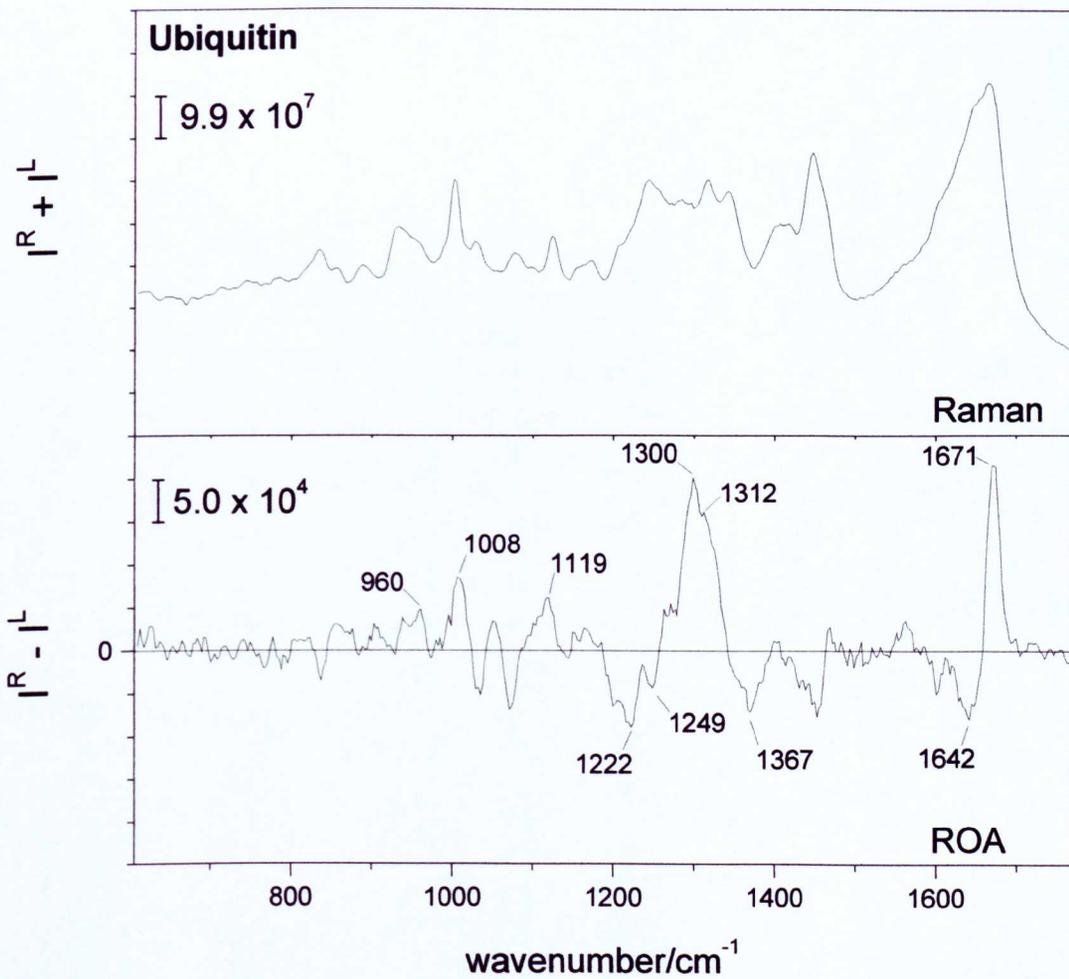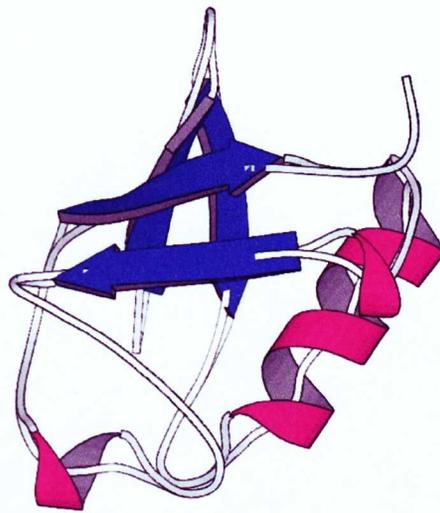
**Figure 4.12.** Molscript diagram and the backscattered Raman and ROA spectra of human ubiquitin in 0.1M acetate buffer with ~8mM KI solution at pH 5.4, 100 mg/mL.

proteins. However, the $\alpha + \beta$ regions are segregated, and this may have an effect on how much influence the $\alpha$-helix content is having on the otherwise strongly $\beta$ signal in the amide I region. The other main signatures of $\beta$ type structure are also clearly present, with well-defined bands at ~1222, ~1249 and 1367 cm$^{-1}$, reflecting an even greater similarity to $\beta$- barrel type structures. However, unlike the previous examples, there is a strong peak at ~1300 cm$^{-1}$ generated by hydrophobic $\alpha$-helix, but is fairly broad to the low wavenumber side, with some other positive structure coming in at around ~1264 cm$^{-1}$ that may be associated with $\beta$-turns. The skeletal backbone stretch region also has definite structure, notably a peak at ~960 cm$^{-1}$, frequently seen in $\alpha + \beta$ proteins. The band at ~1008 cm$^{-1}$ is probably due to calcite from the optics of the experiment, and a clear band at similar wavenumber is seen in the parent Raman spectrum.

### 4.3.4 Summary of mainly beta proteins

Proteins that are classed as mainly beta can have significant amounts of other types of structure that greatly affects their tertiary fold. As seen in the spectra presented in this section, however, ROA is capable of providing remarkably consistent data when presented with a number of proteins that are known to have similar structural properties. As the proteins become more varied in structure, so do their ROA spectra. This is demonstrated by the comparison of the similarity of band positions and relative intensities over the whole set of experimental data, which will be discussed in more detail in Chapter 7. The six mainly $\beta$ protein ROA spectra presented in this section represent the best quality data on this type of molecule to date, and have served to solidify some critical $\beta$-sheet, strand and turn assignments.

### 4.4 Conclusion

The results presented in this chapter represent the largest collection of protein ROA spectra of proteins with similar structural and fold characteristics, and as such contains the most useful information about how subtle differences in structure embody themselves in ROA spectral band patterns. As will be discussed in Chapter 8, the visual similarity of

these proteins can be put into a formal context by means of principal component analysis, but for the purposes of this discussion, simple visual inspection and correlation to the ROA band assignment table shown in Table 2.3 have been used to infer structural information from these spectra. The main difficulties associated with the interpretation of these spectra lies in the variety of conformations that individual β-strands and whole β-sheets may adopt (Salemme, 1983; Nesloney & Kelly, 1996). Coupled with the fact that there are no straightforward model β-structures available to use as a starting reference point (unlike α-helix where α-helical homopolypeptides have been useful), the task of spectral analysis of beta proteins has been all the more difficult. However, as Figure 4.6 demonstrates, the spectra of four proteins with known fold similarities are sufficiently similar to each (and different from the others) to provide a surrogate model for at least β-barrel structures. This being the case, band assignments for beta structure in other proteins has been made slightly harder, but clear differences between ordered β-sheet and unassociated β-strand are becoming apparent.

Only with a more complete understanding of the mechanisms behind ROA band generation (from a combined theoretical and experimental study for example) will the precise contributions to overall band patterns from different types of beta structures be known. However, the 'inverse' approach of obtaining such spectra first, and attributing peaks to structural features and adjusting our understanding with every new spectrum has (as these spectra should hopefully illustrate) provided some solid band assignments and a useful basis for the analysis of many other (more complicated) protein structures.

# Chapter Five
# Alpha Beta Proteins

## 5.1  Introduction

Some of the most complex protein folds belong to the alpha beta ($\alpha\beta$) class of proteins. Members of this structural class owe their amazing diversity in shape and function to the fact that their structures are not dominated by either $\alpha$-helix or $\beta$-sheet, although both structural types are present to a significant extent. The normal definition of an $\alpha\beta$ class protein is one which contains at least ~15% of both $\alpha$-helix and $\beta$-sheet, with similar proportions of each structural element in any given example. However, there are two distinct types of $\alpha\beta$ protein, which are defined by the nature of their $\beta$-sheet component. $\beta$-sheet may exist in either parallel or anti-parallel form, the latter being much more common than the former. An $\alpha\beta$ protein that contains anti-parallel $\beta$-sheet is known as an $\alpha+\beta$ protein, whereas an $\alpha\beta$ protein which contains only parallel $\beta$-sheet is known as an $\alpha/\beta$ protein.

The variety of structures that can be constructed from a roughly equal proportion of $\alpha$-helix and $\beta$-sheet (and a very significant turn content) makes these proteins of key interest for ROA study. Among the most striking fold classes of all is the TIM $\alpha/\beta$ barrel, which is one of the nine 'superfold' classes as described in Chapter 2. Two examples of this fold class are presented in this chapter as well as one other $\alpha/\beta$ protein. In addition, a number of $\alpha+\beta$ proteins are shown and discussed.

## 5.2.  Experimental

Samples of rabbit aldolase and, subtilisin Carlsberg were obtained from Fluka. Ribonuclease A and B, hen ovalbumin, hen and turkey ovomucoid and $\alpha$-amylase were all obtained from the Sigma Aldrich Company. All samples were easily soluble in aqueous solution, and prepared at ~50-100 mg/mL in 0.1M acetate buffer at pH 5.4.

## 5.3 Results and discussion

### Rabbit aldolase

Rabbit aldolase is an $\alpha/\beta$ protein which is a member of the TIM $\alpha/\beta$ barrel superfold class. It contains 363 amino acid residues that form 8 individual subunits that pack to form an eight-stranded parallel $\beta$-sheet in the centre. Each subunit also contains a large $\alpha$-helix, which surround the internal $\beta$-sheet, forming the distinctive rosetta of the TIM $\alpha\beta$ barrel. The crystal structure of rabbit aldolase has been solved (Gamblin et al., 1990) and is illustrated in Figure 5.1. as a Molscript diagram. The crystal structure suggests that both the $\alpha$-helical and $\beta$-sheet elements of the protein will be well resolved, as the structural elements themselves are extremely well-defined. In addition, with the $\alpha$-helices being positioned as they are, it is expected that the solution structure of the TIM barrel will reveal significant hydration of the $\alpha$-helices, as well as some sort of signature of amphipathicity.

The backscattered Raman and ROA spectra of aldolase in aqueous solution are shown in Figure 5.1. Interestingly, the immediate impression is one of a protein with a significant $\alpha$-helical content. The amide III region is dominated by strong sharp bands at ~1306 and ~1342 cm$^{-1}$, attributed to hydrophobic and hydrophilic $\alpha$-helix respectively. This correlates well with the expectations from the crystal structure. In addition, the amide I region is very similar to that seen in the alpha proteins. Negative structure at ~1253 cm$^{-1}$ may be attributed to $\beta$-sheet, but the lack of a clear negative peak at ~1215 – 1225 cm$^{-1}$ (see table 2.4) suggests that this is not hydrated $\beta$-sheet, again in agreement with expectation.

Perhaps the most valuable information from this type of protein is the clear lack of any peak at ~1360 cm$^{-1}$. The survey of beta proteins presented earlier reveals a consistent pattern of negative structure in this region and can be seen most prominently in proteins with a high $\beta$-hairpin content. However, these hairpins are absent from the parallel $\beta$-sheet in aldolase, and the ROA spectrum shown here is an excellent example of how ROA can reproducibly detect this subtle feature within such a complex target molecule.
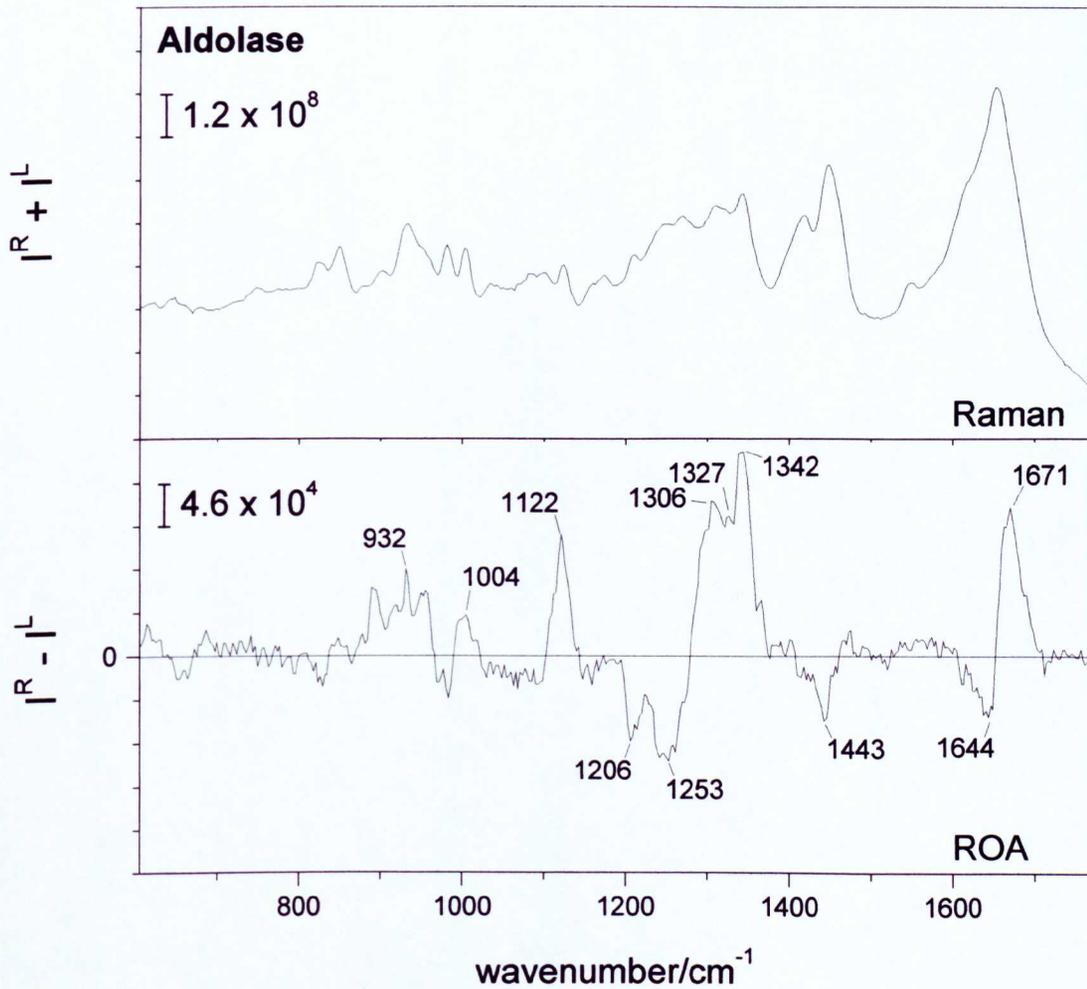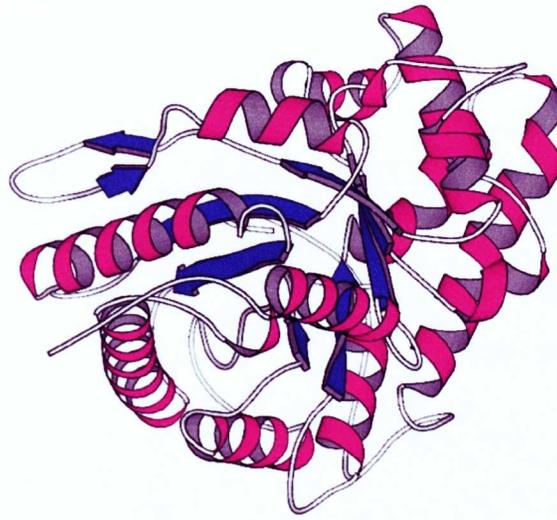
**Figure 5.1** Molscript diagram and the backscattered Raman and ROA spectra of the rabbit aldolase, measured at ~67mg/mL in 0.1M acetate buffer at pH 5.4.

## Bovine α-amylase

Despite their abundance in nature, TIM α/β barrel proteins are not commonly available or perfectly straightforward to deal with. However, bovine α-amylase is a protein that comprises two separate domains, one of which is TIM barrel-like, the other composed of a β-sheet in a Greek-key motif. α-amylase is made up from 476 amino acids, with the TIM α/β barrel domain being of similar size to that seen in aldolase. The crystal structure of α-amylase (Swift et al., 1991) is illustrated in Figure 5.2 as a Molscript diagram. The TIM barrel domain appears similar to that of aldolase, but the additional β-barrel domain can be seen on the left hand side of the diagram. Unlike the uniform ring of solvent exposed α-helices as seen in the previous example, the α-helices in α-amylase are less well-defined due to the proximity of the β-sheet domain. In addition, α-amylase is believed to be one of the few folded proteins in this current study to have a small but significant carbohydrate content. The carbohydrate content of a protein can have a significant effect on the appearance of the ROA spectrum of a protein, but generally bands are weak unless a large amount (~50%) is present.

The backscattered Raman and ROA spectra of bovine α-amylase in aqueous solution are shown in Figure 5.2. It is immediately apparent that the data is of poorer quality than that of previous examples. This is mainly due to experimental factors (e.g. the protein degrades in the laser beam), but may also be due to the significant carbohydrate content of this protein. There is practically no band at all ~1340 cm$^{-1}$ suggesting that the α-helices are not nearly as solvent exposed in this protein as with aldolase. The amide I region however indicates the significant presence of α-helix. Once again, there is no significant β-hairpin signature at ~1360 cm$^{-1}$, but a clear peak at ~1255 cm$^{-1}$ is coming from β-sheet. The other main feature of this spectrum is the overall pattern of bands, which is on closer examination quite similar to that of aldolase. Similar in terms of band position, three areas of positive structure at ~900, ~1050 and ~1144 cm$^{-1}$ are all suggestive of a mixed alpha and beta structure, however the detail in this region (especially relative band intensities) is deemed to be too unreliable to be of much usefulness in this case due to the presence of contributions from carbohydrate.
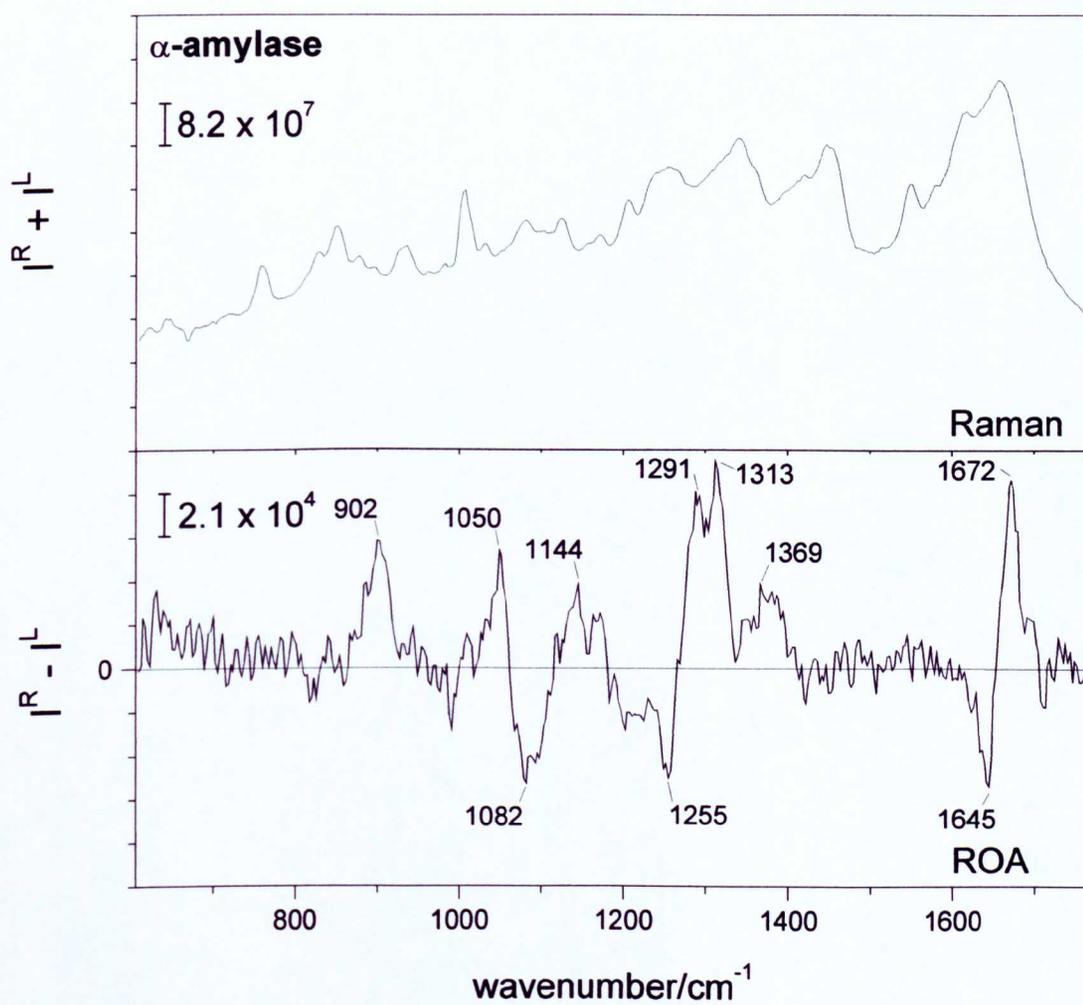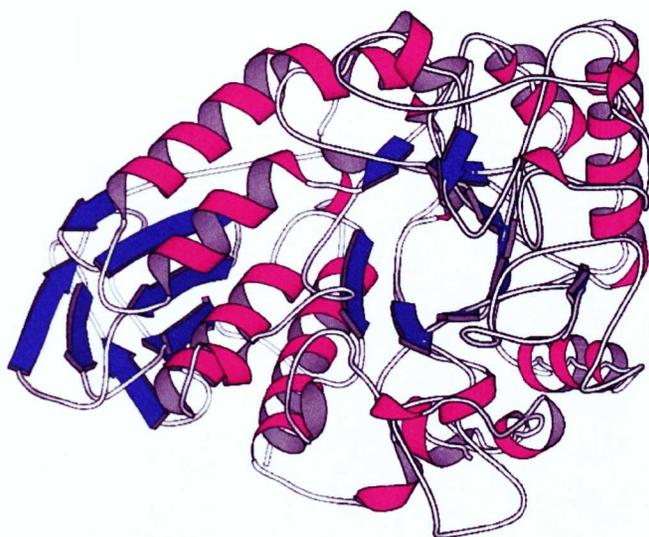
**Figure 5.2** Molscript diagram and backscattered Raman and ROA spectra of bovine α-amylase measured at ~100mg/mL in 0.1M acetate buffer at pH 5.4.

## Subtilisin Carlsberg

Bacterial protease subtilisin Carlsberg is another $\alpha/\beta$ protein, but unlike aldolase is not a TIM $\alpha/\beta$ barrel. The structure of this protein is described as a three-layered alpha-beta-alpha sandwich (Neidhart et al., 1988), and is illustrated in Figure 5.3. Containing ~30% $\alpha$-helix and 18% $\beta$-strand, this protein is a borderline case for mainly alpha classification (as is aldolase), but the mixed connectivity of the alpha and beta components, and the overall $\alpha/\beta$ structure mean that this protein is structurally quite distinct from the mainly alpha proteins such as the lysozymes.

The backscattered Raman and ROA spectra of bacterial protease subtilisin Carlsberg is shown in Figure 5.3. The ROA spectrum is dominated by the strong positive band centered at ~1306 cm$^{-1}$ from hydrophobic $\alpha$-helix, with only a shoulder at ~1342 cm$^{-1}$ from hydrated $\alpha$-helix. Positive structure at ~942 and ~1123 cm$^{-1}$ is similar to that seen in proteins containing $\alpha$-helix, and the amide I region also suggests a mainly alpha protein. However, the rest of the detail in the spectrum precludes this protein from such a classification. Most significantly, there is absolutely no structure at ~1360 cm$^{-1}$ from beta-hairpins which, along with aldolase, provides a unique indicator that the beta character of this protein is significantly different to that usually observed. As well as this feature, another interesting aspect of this spectrum is the broadening (to the low wavenumber side) of the $\alpha$-helix band peaking at ~1306 cm$^{-1}$. This is also slightly apparent in the same peak in the spectrum of aldolase, and is possibly due to the significant turn content of the $\alpha/\beta$ proteins in generally.

The ROA spectrum of subtilisin Carlsberg also has very well-defined $\beta$-structure, with peaks at ~1220 cm$^{-1}$, ~1242 cm$^{-1}$ and 1266 cm$^{-1}$, indicating a possible mixture of hydrated and non-hydrated $\beta$-sheet. Overall, this spectrum represents one of the best examples of a mixture of alpha and beta signatures in one protein, as well as a distinctly different type of tertiary fold, even from the other examples of proteins of the same structural class such as aldolase and $\alpha$-amylase.
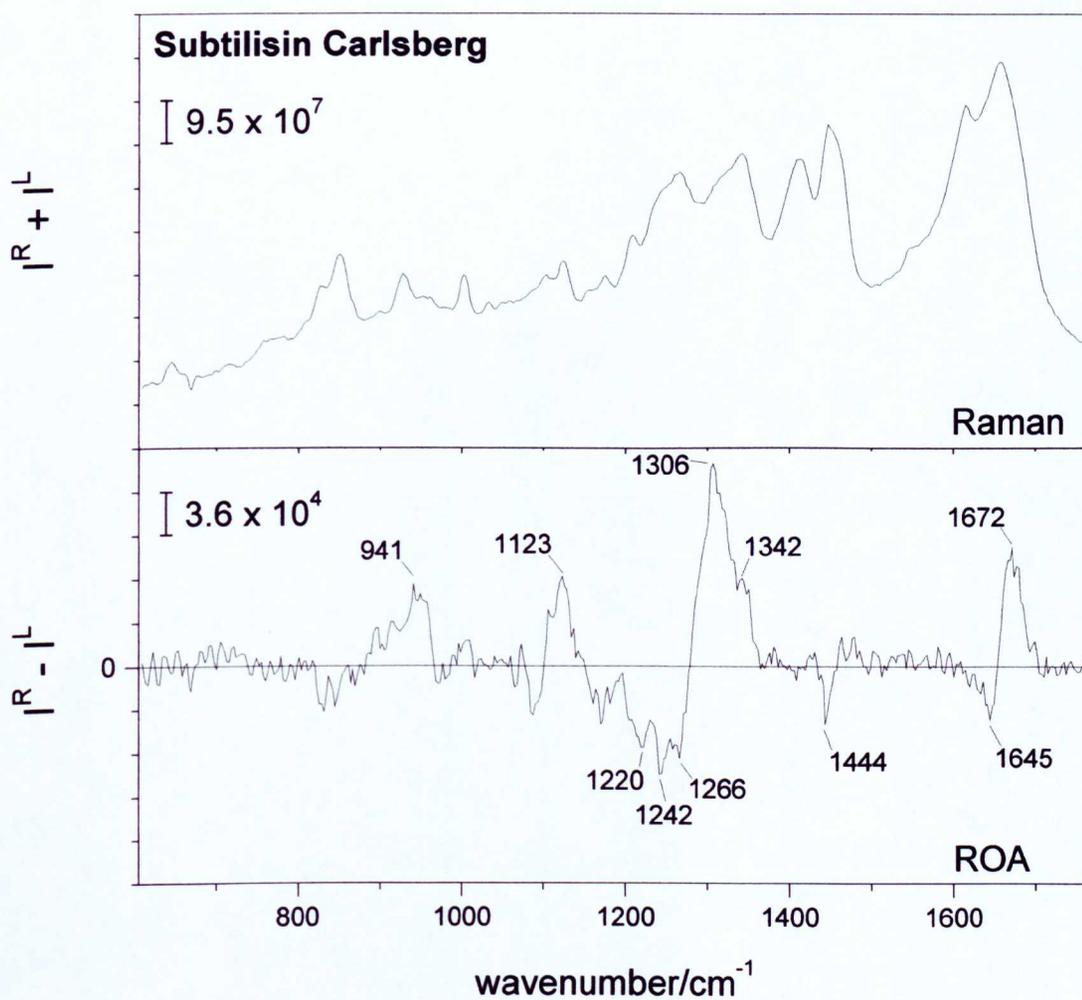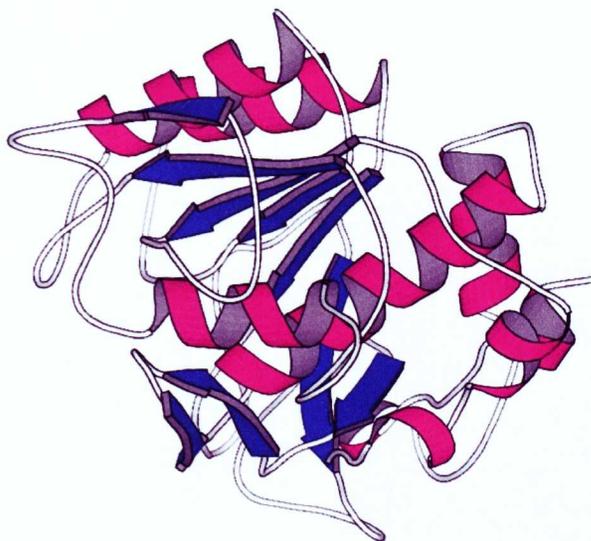
**Figure 5.3** Molscript diagram and backscattered Raman and ROA spectra of bacterial subtilisin Carlsberg measured at ~100 mg/mL in 0.1M acetate buffer at pH 5.4.

## Hen ovalbumin

Hen ovalbumin has a structure like those of the serine proteinase inhibitors or 'serpins', a family of proteins that possess an active loop which is capable of docking with its substrate to form a stable proteinase-proteinase inhibitor complex (Perutz, 1992). Hen ovalbumin however is inactive (Perutz, 1992) and its function is as yet unclear. It is a large protein made up from 385 amino acids, and is rich in secondary structure with ~30% α-helix and ~32% β-sheet. Its crystal structure has been solved (Stein et al., 1991) and is illustrated in Figure 5.4. The arrangement of the β-strands in hen ovalbumin is decidedly more complex than in other proteins in this section and is a mixture of parallel and anti-parallel β-sheet. Although there is parallel β-sheet present, it is not an α/β protein as the majority of β-sheet is anti-parallel.

The backscattered Raman and ROA spectra of hen ovalbumin in aqueous solution are shown in Figure 5.4. There is no negative peak at ~1360 cm$^{-1}$, which suggests few (if any) hairpin bends, and may also be a result of strong positive structure at ~1342 cm$^{-1}$ counteracting the negative ~1360 cm$^{-1}$ band. α-helical signatures are evident in abundance with peaks at ~ 927 and ~1342 cm$^{-1}$. The amide I region is indicative of a protein with a high proportion of α-helix with the usual negative-positive couplet peaking at ~1671 cm$^{-1}$, although broadening out slightly to higher wavenumber due to the significant contribution from β-sheet. Further evidence of the high β-content of hen ovalbumin is given by the strong and sharp negative ROA band centered at ~1218 cm$^{-1}$, suggesting that the β-strand content may be hydrated. However, the main feature of this spectrum is the strong positive band peaking at ~1312 cm$^{-1}$. This band is the familiar band attributed to hydrophobic α-helix as seen in several examples already, which usually peaks at between 1295-1305 cm$^{-1}$, but has been shifted slightly to ~1312 cm$^{-1}$ which suggests that a contribution is being made by the plentiful loop and turn structure, which accounts for positive structure between ~1315-1320 cm$^{-1}$.

The overall pattern of ROA bands observed in this spectrum is unequivocally that of an αβ protein. The fold characteristics of this protein are different to any other in the data set, a fact that is reflected by the individual character of the spectrum, yet

the main peaks all correspond well to established ROA band assignments, being consist with information from the crystal structure.
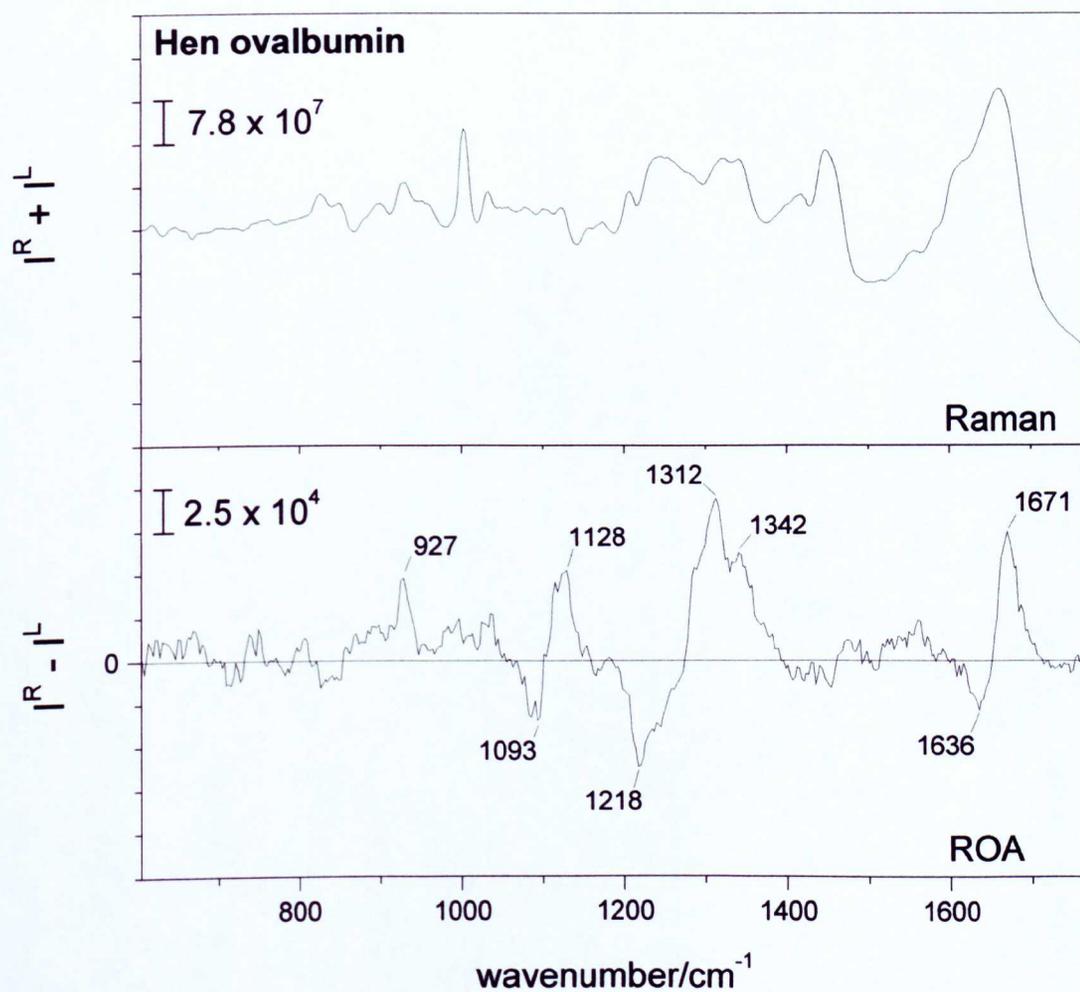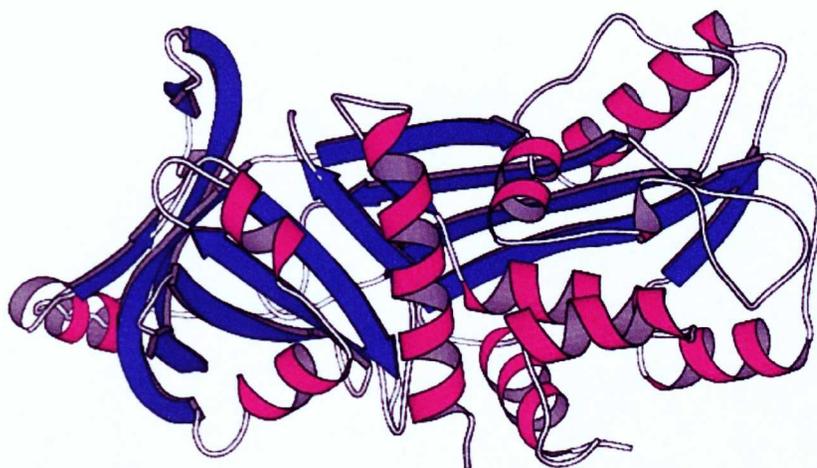


**Figure 5.4** Molscript diagram and backscattered Raman and ROA spectra of hen ovalbumin, measured at ~100 mg/mL in 0.1M acetate buffer at pH 5.4.

**Hen and turkey ovomucoid**

These proteins are serpins and are glycoproteins with a carbohydrate content of ~25% (Watanabe et al., 1981). They comprise three domains each of ~56 residues and the overall protein unit contains ~18% α-helix and ~16% β-strand. The complete crystal structures for both of these proteins are unavailable, however one domain of silver pheasant ovomucoid is shown in Figure 5.5. Structural studies of the ovomucoids reveal that they are all very similar, with a variable amount of β-strand ranging from ~10-21 % (Weber et al., 1981; Hoogstraten et al., 1995). The structure is dominated by a single β-sheet that is formed from the β-strand in each domain, however they are α+β proteins.
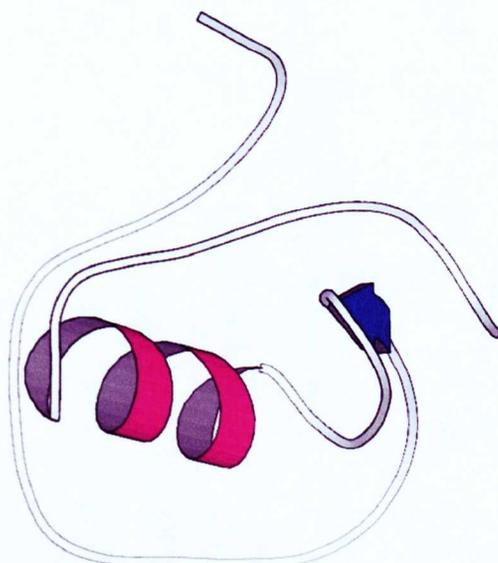


**Figure 5.5** Molscript diagram of silver pheasant ovomucoid (PDB code 2ovo, Bode et al., 1985).

The backscattered Raman and ROA spectra of both hen and turkey ovomucoid are shown in Figure 5.6. Although they are similar, there are sufficient differences in fine detail between the two spectra to qualify the statement that they are quiet distinct samples. The difference in terms of structure may be slightly exaggerated in that, even from one species to another, the structures are very similar and hence the ROA spectra should reflect that. It should be noted that both samples were prepared and analyzed under the same experimental conditions, and yet the quality of the hen ovomucoid spectrum is clearly superior to turkey ovomucoid. However, the exact purity of each sample is not known, and the behavior of each sample in the laser beam is sometimes greatly affected by factors within the sample that are not related to the structure of the
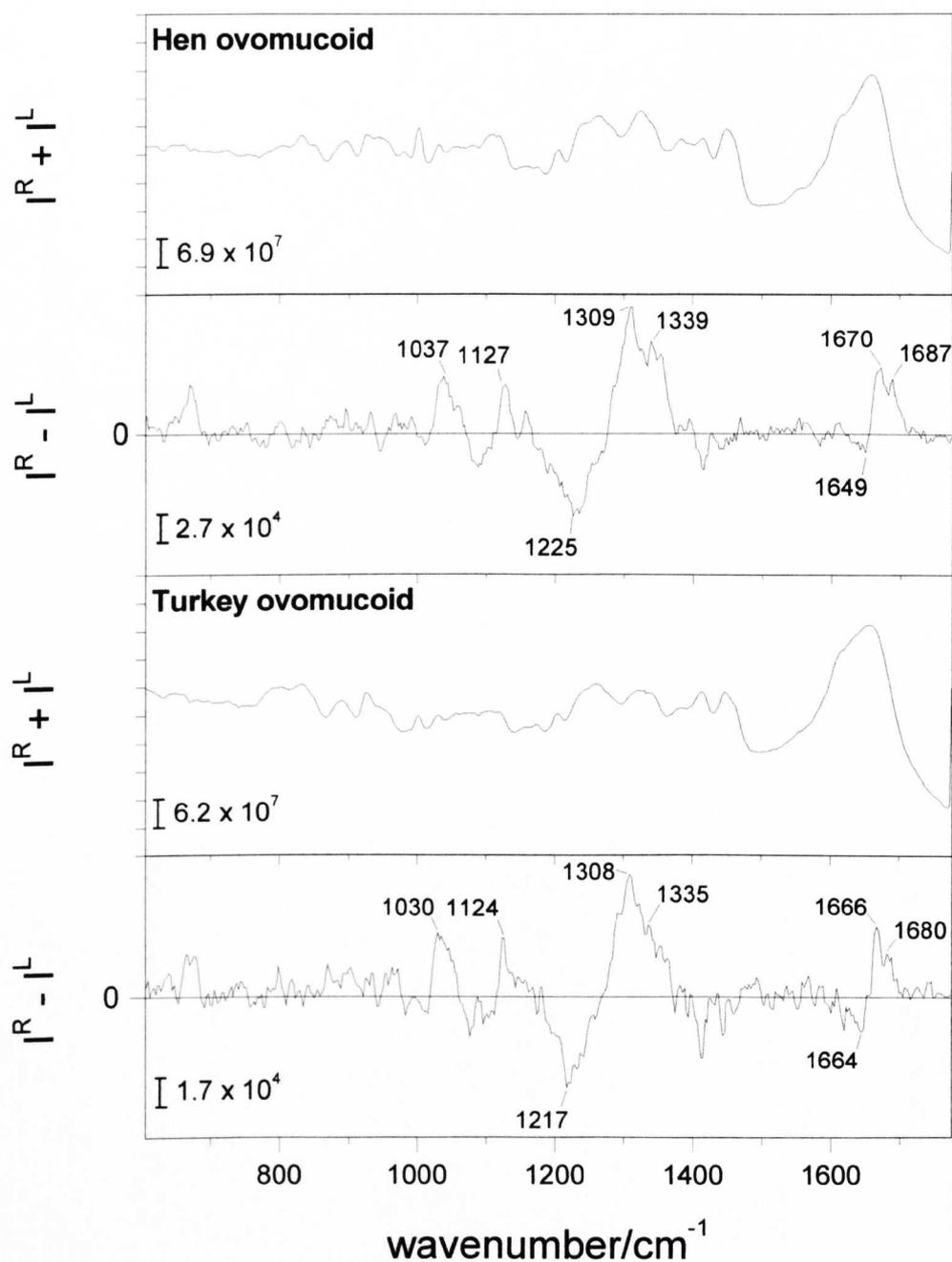
**Figure 5.6** The backscattered Raman and ROA spectra of hen (top pair) and turkey (bottom pair) ovomucoid at ~100mg/mL in 0.1M acetate buffer at pH 5.4.

proteins themselves. To this end, this example of two very similar proteins is useful in giving some insight into the reliability of ROA spectroscopy as a structure elucidation technique, and how the quality of the original samples is very important in determining the quality of the experimental data. The overall appearance of the ROA spectrum of hen ovomucoid (with the exception of the amide I region) is very similar

to that of hen ovalbumin, with turkey ovomucoid being slightly less well-defined. The amide I region is possibly the most revealing part of these spectra however. Each has two distinct peaks, one corresponding to an α-helical assignment (~1666 - 1670 cm$^{-1}$) and also a higher peak corresponding to both β-sheet and loop and turn structure (~1680 - 1687 cm$^{-1}$). Both amide I regions are of medium strength, and their negative components (at lower wavenumber) are weak, reflecting the relatively low total amount of secondary structure. They are also quite broad, suggesting that loop and turn structure is a dominant feature.

Otherwise, both spectra share common characteristics that reflect their dual alpha and beta nature, principally the strong positive bands in the amide III region corresponding to both hydrophobic and hydrophilic α-helix, and a negative band at ~1225 and ~1217 cm$^{-1}$ respectively, coming from β-strand.

### Bovine pancreas ribonucleases A and B

The ribonucleases are digestive enzymes secreted by the pancreas and are α+β proteins of ~124 amino acids. Ribonuclease A is non-glycosylated while ribonuclease B is glycosylated with ~12% carbohydrate (Rudd et al., 1994). Crystal structures have
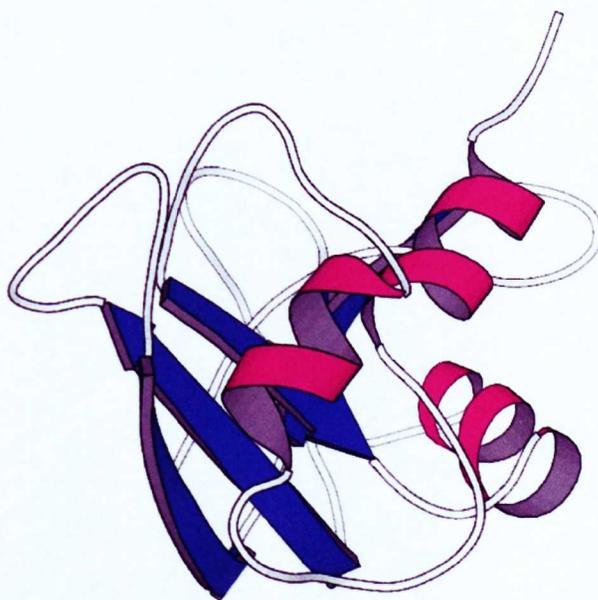


**Figure 5.7** Molscript diagram illustrating the crystal structure of ribonuclease A (PDB code 7rsa).

been solved for both proteins (PDB codes 7rsa and 1rbb respectively) and reveal very similar structures. The crystal structure of ribonuclease A is shown in Figure 5.7 and is a useful guide to the structure of both proteins. Both contain ~18% α-helix and ~33% β-strand which form two anti-parallel β-sheets. They arrange to form an αβ roll
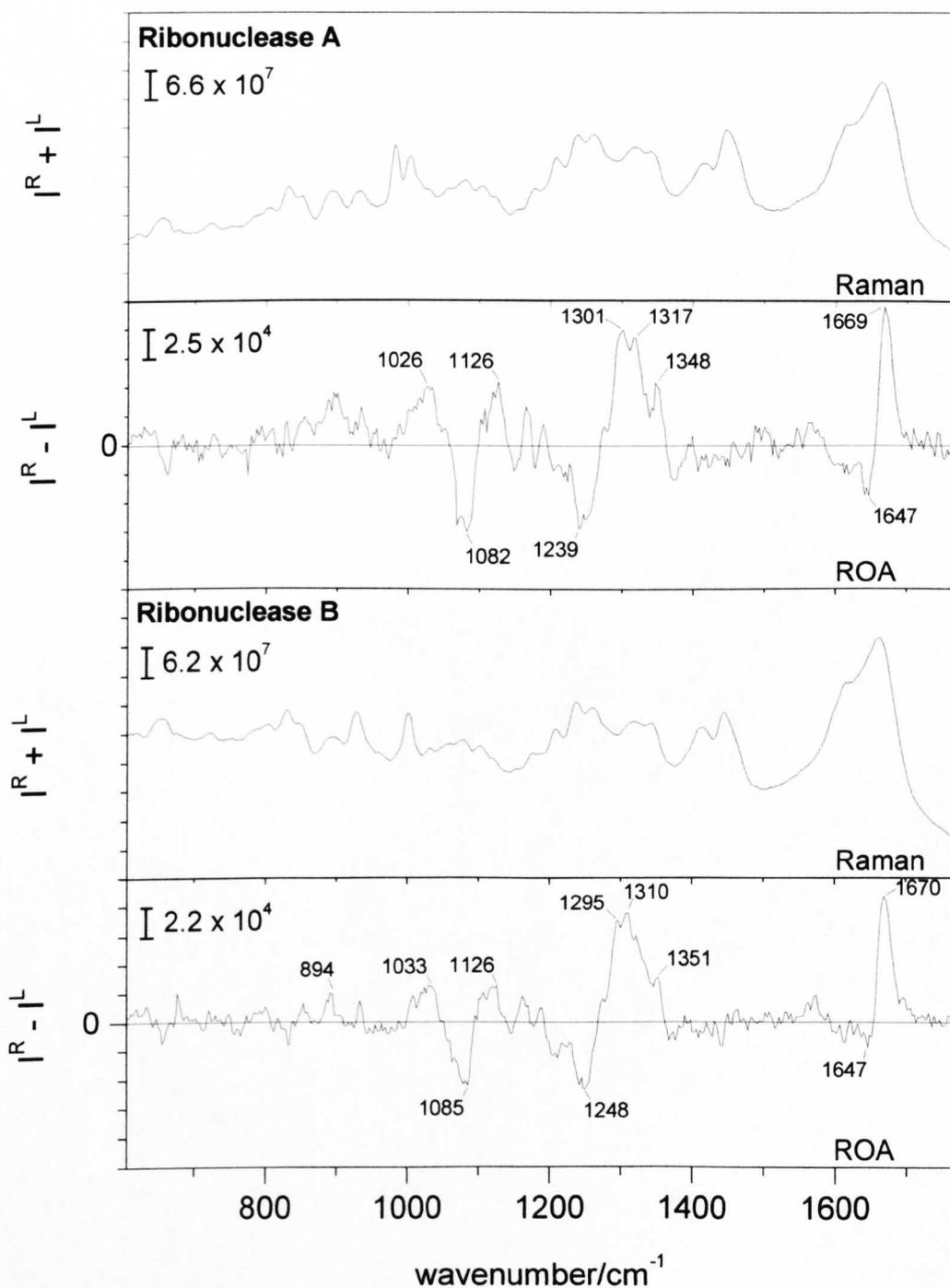


**Figure 5.8** The backscattered Raman and ROA spectra of ribonuclease A (top pair) and ribonuclease B (bottom pair) measured at ~100 mg/Ml in 0.1M acetate buffer at pH 5.4.

structure with a curved β-sheet (Branden & Tooze, 1991). The backscattered Raman and ROA spectra of ribonucleases A and B are shown in Figure 5.8. On initial inspection, they are very similar, but again their subtle differences. Again, this is possibly due to the fact that each sample is prepared independently and relative purities are not known. On further inspection, the overall appearance of each spectrum reveals a much higher degree of similarity to the mainly beta proteins than any others seen in this section. The extremely strong, sharp amide I peak, despite peaking at ~1670 cm⁻¹ in each case, is similar to that seen in the all beta proteins such as hen avidin. Strong negative structure in the ~1238-1245 cm⁻¹ region also strongly supports the mainly beta assignment. However, the α+β character of these proteins is clearly apparent from the positive structure in the amide III region once again. The negative band at ~1085 cm⁻¹ is a unique feature of these spectra and has been assigned tentatively to right-twist β-strand, complementing the observations from the crystal structure that there is curved β-sheet present.

## 5.4 Conclusion

The variety of possible structures that can be generated by the combination of α-helix and β-strand/sheet are virtually limitless, and the several examples in this chapter barely scratch the surface. However, the point of this survey has been to reveal how proteins that are far from model structures can yield a mass of structural information from their ROA spectra, despite differences in sample quality. The most interesting fact relating to the spectra presented in this chapter is the ability of ROA to consistently identify structure that is known to be present, and is presented against the backdrop of the tertiary fold of the protein as well. These examples demonstrate the fact that ROA gives more than just a guess at whether certain types of structure are present, but a realistic opportunity to gain quantitative measures of individual structural types and fold classes. They also serve to illustrate the amount of work that remains to be done to achieve the goal of quantitative spectral analysis. As will be discussed later, it is still very difficult to describe the data subjectively, especially when each protein has such individual character and structural complexity. Nevertheless, these spectra represent a unique insight into the 'grey area' where most other non-high resolution techniques struggle to yield any useful information at all.

# Chapter Six
# Proteins with Irregular Folds

## 6.1 Introduction

The task of making band assignments in the ROA spectra of proteins has been investigated thoroughly by correlating experimental observations with the crystal structures of proteins for which they are available. However, crystal structures have only been solved for a comparatively small number of proteins. The reason for this is that solving the crystal structure of a protein is by no means a straightforward process. In fact, many proteins will simply not crystallize due to their lack of any structural regularity (amongst other things). The study of this type of protein must therefore depend upon the myriad of other physico-chemical techniques that are able to shed some light on the nature of protein structure. To this end, ROA has been used to look at a number of such proteins where little (if any) information pertaining to secondary structure content and tertiary/quaternary arrangement exists.

Far from being an area of secondary interest, the study of proteins with irregular folds is of key importance in determining how the less well established band assignments pertaining to sections of random coil, PPII helix and the various loops and turn structures figure in the ROA spectra of all types of proteins.

In addition to simply exploring the relationship between bands in the ROA spectra of proteins with irregular folds and the structure of the proteins themselves, the spectra presented in this chapter have allowed us to gain some insight into the different types of disorder present. (Smyth et al., 2001)

## 6.2 Experimental

Lyophilized powder samples of baker's yeast invertase (Grade VII), soybean Bowman-Birk inhibitor and bovine orosomucoid ($\alpha_1$-acid glycoprotein) were supplied by the Sigma-Aldrich Company Ltd., and no further purification was required before analysis. A sample of $Cd_7$-metallothionein (from rabbit liver) in Tris buffer solution

was supplied by Milan Vašák of the University of Zurich. A sample of anti-freeze protein III (AFP III) was supplied by Unilever.

## 6.3 Results and discussion

The following Raman and ROA spectra were collected as part of a larger study into the solution structure of native proteins with irregular folds, as well as unfolded proteins and disordered polypeptides. (Smyth et al., 2001) Therefore, as part of the discussion of these results, some spectra of polypeptides (poly-L-glutamic acid and poly-L-lysine) are referred to, although the author did not collect these spectra.

### Yeast invertase

Yeast external invertase is a biocatalyst used in the sugar industry. It is a globular protein with a molecular weight of ~270 kDa encompassing two subunits with up to 50% carbohydrate by mass in the form of nine high-mannose oligosaccharide chains, each linked to an asparagines residue on the polypeptide backbone. There exists very little information on the structure of this protein, with the small amount that is available appearing quite contradictory. There is no crystal structure for invertase at the present time. UVCD studies suggest little or no α-helix (Schülke et al., 1988) and possibly a significant amount of β-sheet, perhaps in the form of a β-propeller fold (Pons et al., 1998). On the other hand, a conventional Raman study suggests a mainly α-helical structure in aqueous solution, and β-sheet in the solid state (Athès et al., 1998).

The backscattered Raman and ROA spectra of yeast invertase in aqueous solution are shown in Figure 6.1. The lack of a negative-positive couplet in the amide I region once again suggests the presence of little or no α-helix and β-sheet. In fact, there is only a very weak peak at 1673 cm$^{-1}$. However, there are negative peaks within the amide III region at ~1220 and 1250 cm$^{-1}$ that suggest the presence of some type of β-sheet. This immediately seems to contradict the conclusion drawn from the absence of a couplet in the amide I region, but may be explained as the result of the presence of amino acids within β-strands which do not hydrogen bond to form β-sheet which

would be reflected in the amide I region, since interstrand bonding involves the C = O group. The presence of unassociated β-strands may be supported by the high carbohydrate content (~50%). In addition, the main positive peak in the amide III region at ~1312 cm$^{-1}$ may originate in turn structure (Smyth et al., 2001).
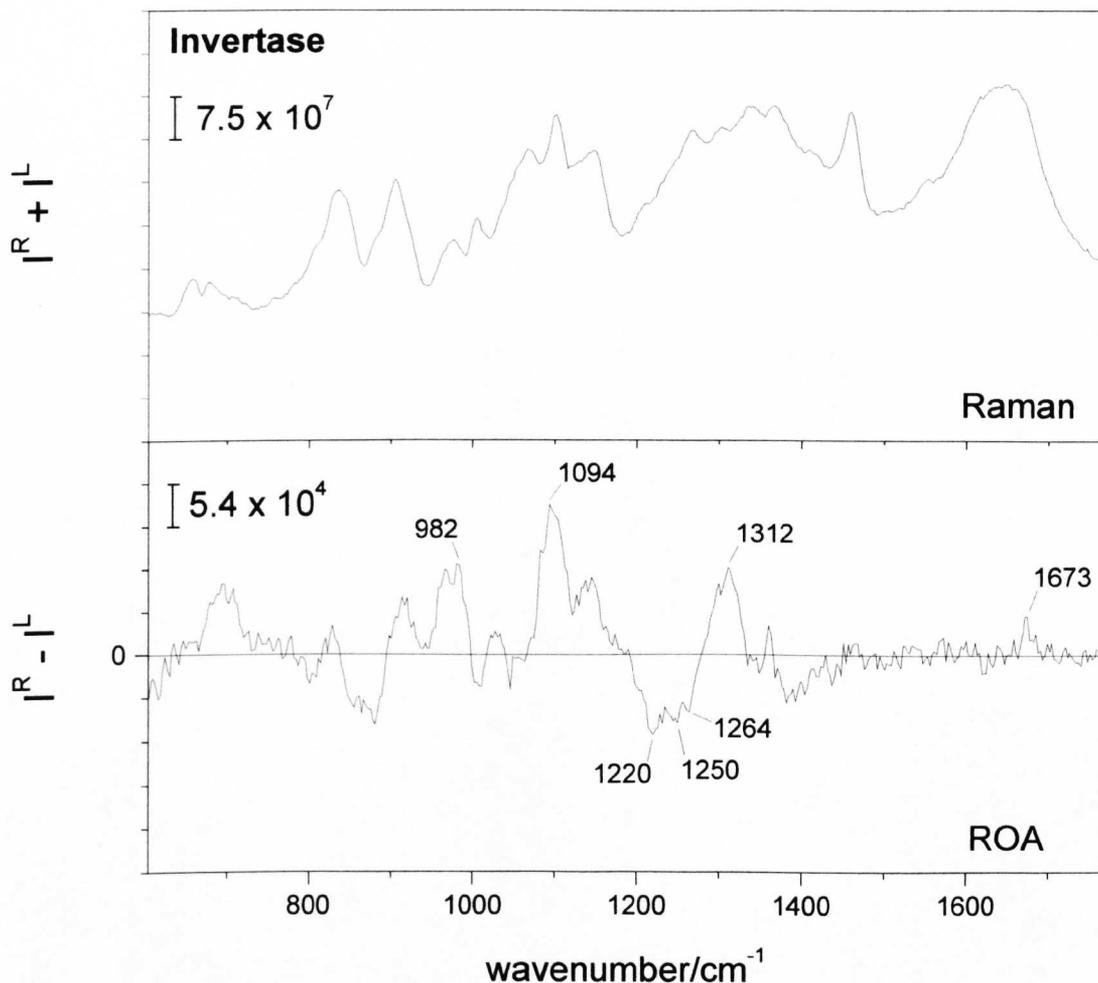


**Figure 6.1** The backscattered Raman and ROA spectra of yeast external invertase in 0.1M acetate buffer with ~8mM KI at pH 5.4, at ~50 mg/mL.

However, the most prominent features within the ROA spectrum of yeast external invertase appears at below ~1150 cm$^{-1}$ and are most likely attributable to the mannose-based oligosaccharide chains. The prominent ROA band at ~1094 cm$^{-1}$ may originate in the vibrations of peripheral phosphate groups in oligosidic chains since they are known to generate conventional Raman bands at 1081 cm$^{-1}$ (Athès et al., 1998).

# Bowman-Birk protease inhibitor

Bowman-Birk (trypsin/chymotrypsin) protease inhibitors are small proteins with a molecular weight ~7-9 kDa. They are found in the seeds of many leguminous plants, for example soybean, from which this sample was derived. They are characterized by the unusually high number of seven disulfide links that hold this small polypeptide chain in a well-defined native fold. These proteins are of considerable interest for medical and agricultural applications (de la Sierra et al., 1999) due to their antiprotease activity, anticarcinogenic activity, immune-stimulating properties, and their ability to defend plants from insect attack (Smyth et al., 2001).

The structure of soybean Bowman-Birk protease inhibitor is well defined, and the crystal structure of the PI-II isoform (61 residues) has been resolved (PDB code 1pi2) (Chen et al., 1992). The solution NMR structure of the BBI-I isoform (71 residues) is also available (PDB code 1bbi, Werner et al., 1992) and is shown as a Molscript diagram in Figure 5.3. According to the CATH classification system (Orengo et al., 1997), the Bowman-Birk inhibitors have an irregular architecture with few secondary structures. As can be seen from the solution NMR structure illustrated in Figure 6.2, the protein is a single chain comprising two tandem homologous domains, each containing a small anti-parallel β-sheet which act as protease binding sites.

The backscattered Raman and ROA spectra of the soybean Bowman-Birk protease inhibitor are also shown in Figure 6.2. It contains sharp detailed structure throughout, more similar to proteins with regular folds. Indeed, structural studies including the aforementioned solution NMR structure reveal ~25% β-sheet, whereas a solution UVCD study suggests as much as 41% β-sheet, 28% random coil and 9% α-helix (de la Sierra et al., 1999). Prominent spectral features include a strong sharp negative peak at ~1239 cm$^{-1}$, consistent with a significant content of well-defined β-strand, and strong positive band at ~1322 cm$^{-1}$ suggesting a large amount of residues with PPII structure within the long loops that are known to be present. These two points are consistent with the Ramachandran plot for the protein.
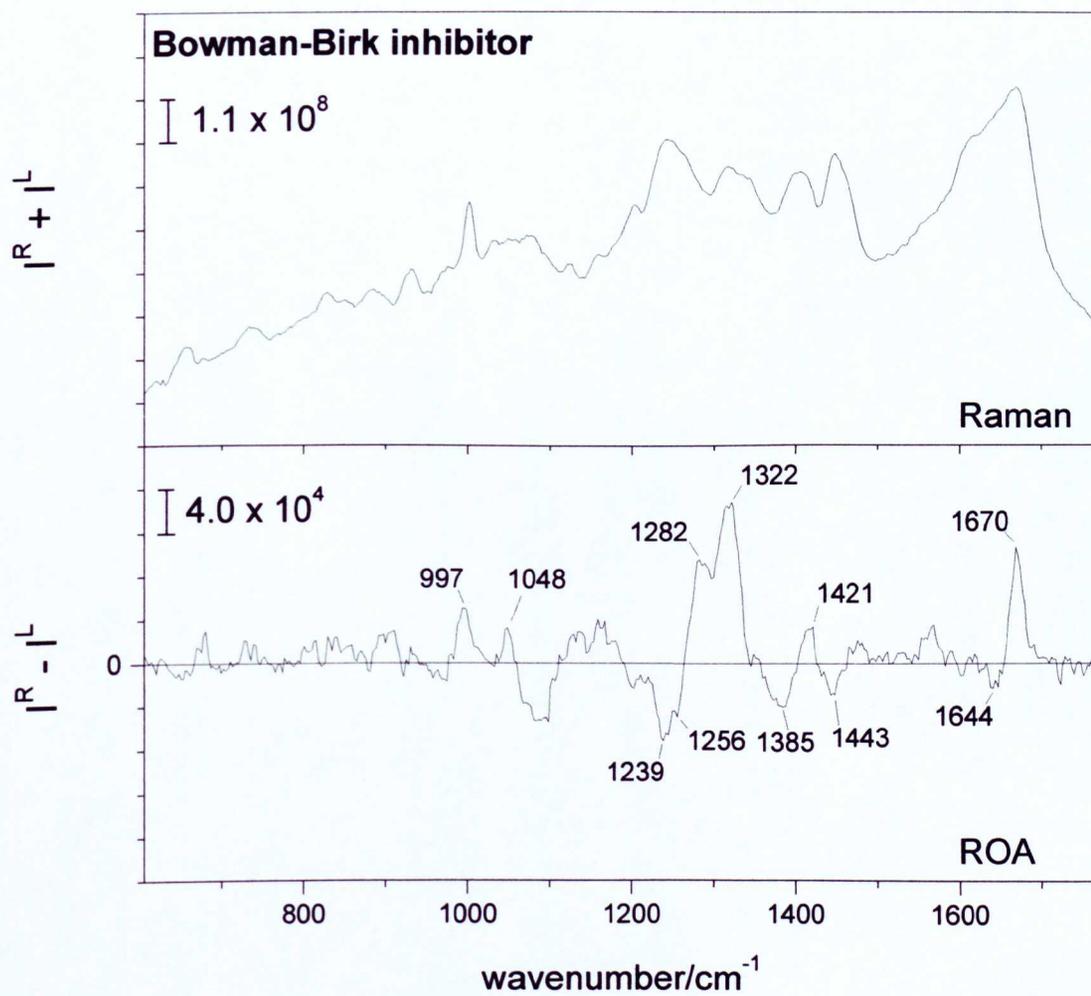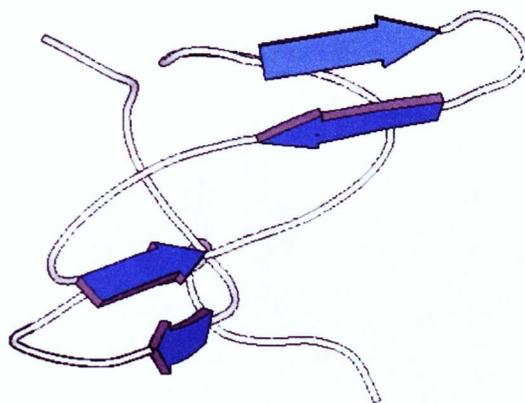
**Figure 6.2** Molscript diagram and backscattered Raman and ROA spectra of Bowman-Birk protease inhibitor in 0.05M acetate buffer at pH 5.4 at 100 mg/ml.

The negative band centered at ~1385 cm$^{-1}$ has been assigned to hairpin bends and is seen in a number of proteins with a lot of β-strand and sheet, and is consistent with the solution NMR structure, 1bbi, which indicates the presence of two such bends.

As well as the amide III region, the sharp couplet in the amide I region tending to higher wavenumber suggests the presence of significant β-sheet similar to the β-barrel proteins discussed in Chapter Four. In summary, the ROA spectrum of soybean Bowman-Birk inhibitor is consistent with the description of a constrained structure containing residues with φ,ψ angles in well-defined regions (β and PPII) of the Ramachandran surface (Smyth et al., 2001).

**Rabbit metallothionein**

Metallothioneins are small cysteine-rich proteins with molecular weights of ~6 kDa. They have the ability to accommodate heavy metal ions, as in this sample which contains 7 Cd$^{2+}$ ions. They also contain very few aromatic residues; in fact, there are no aromatic residues present in the amino acid sequence of 61 residues of this particular protein. Instead of forming disulfide bridges, the 20 cysteine residues bind the divalent metal ions that serves to stabilize the native fold of the protein (Braun et al., 1992).

The X-ray crystal structure of rat metallothionein is available (PDB code 4mt2, Robbins et al., 1991) and is illustrated below in Figure 6.3 Also, solution NMR (Schulze et al., 1988) and conventional Raman (Pande et al., 1986) studies of metallothioneins point to a largely turn-containing structure, with little or no secondary structure, with a possibility of some 3$_{10}$-helix (Messerle et al., 1990).

Figure 6.3 shows the backscattered Raman and ROA spectra of rabbit Cd$_7$-metallothionein. Again, the ROA spectrum contains sharp detail throughout, but with noticeably fewer peaks than in the Bowman-Birk spectrum, which suggests fewer distinct structural types (Smyth et al., 2001). The strong evidence for β-type structures, as seen in the Bowman-Birk inhibitor, are absent from the ROA spectrum of rabbit metallothionein. Specifically, there are no negative bands in the amide III

region between ~1220-1260 cm$^{-1}$, and no negative-positive couplet or strong sharp positive peak at higher wavenumber (1670-1680 cm$^{-1}$) in the amide I region.

The main peak in the spectrum is the strong positive band in the amide III region centered at ~1313 cm$^{-1}$, with a broad shoulder extending down to ~1275 cm$^{-1}$ to the low wavenumber side. This is possibly a reflection of the large amount of residues with $\phi,\psi$ angles within the general $\alpha$ region of the Ramachandran surface. Indeed, this region of the Ramachandran surface for crystal structure 4mt2 (the rat variant of metallothionein) is significantly populated and includes such types of structure as $3_{10}$-helix and type I $\beta$-turns. The assignment of this particular feature to general $\alpha$ type residues as opposed to helical structure is reinforced by the absence of any well-defined helical bands elsewhere in the spectrum. In particular, lack of a couplet (tending to lower wavenumber at ~1660-1670 cm$^{-1}$) in the amide I region support the data from the crystal structure which suggests no $\alpha$-helix is present. The ROA couplet in the amide I region is particularly sensitive to the intrachain hydrogen bonds required to form regular $\alpha$-helical structure, which is a similar conclusion to the presence of unassociated $\beta$-strands in yeast invertase discussed earlier in this chapter.

Another interesting feature within this ROA spectrum is the strong sharp negative band at ~1201 cm$^{-1}$. Usually there is very little ROA intensity in this particular region. It is thought that this peak may reflect the high proportion (~30% of residues) of cysteines in this protein. This assignment is supported by the observation of a similar band in the ROA spectrum of a concentrated solution (~100 mg/mL) of the tripeptide glutathione ($\gamma$-Glu-Cys-Glu) in H$_2$O and also in that of cysteine itself (Gargaro et al., 1993).

In summary, the absence of all $\beta$-type structures like those seen in the Bowman-Birk inhibitor is probably due to the complete absence of aromatic residues, which are known to increase the population of $\beta$-space (Smith et al., 1996). As can been seen from the PDB structure 4mt2 for rat metallothionein, the Ramachandran surface indicates only a small number of residues occupying the general $\beta$ region. However, the overall appearance of the ROA spectrum reflects a constrained but well-defined structure.
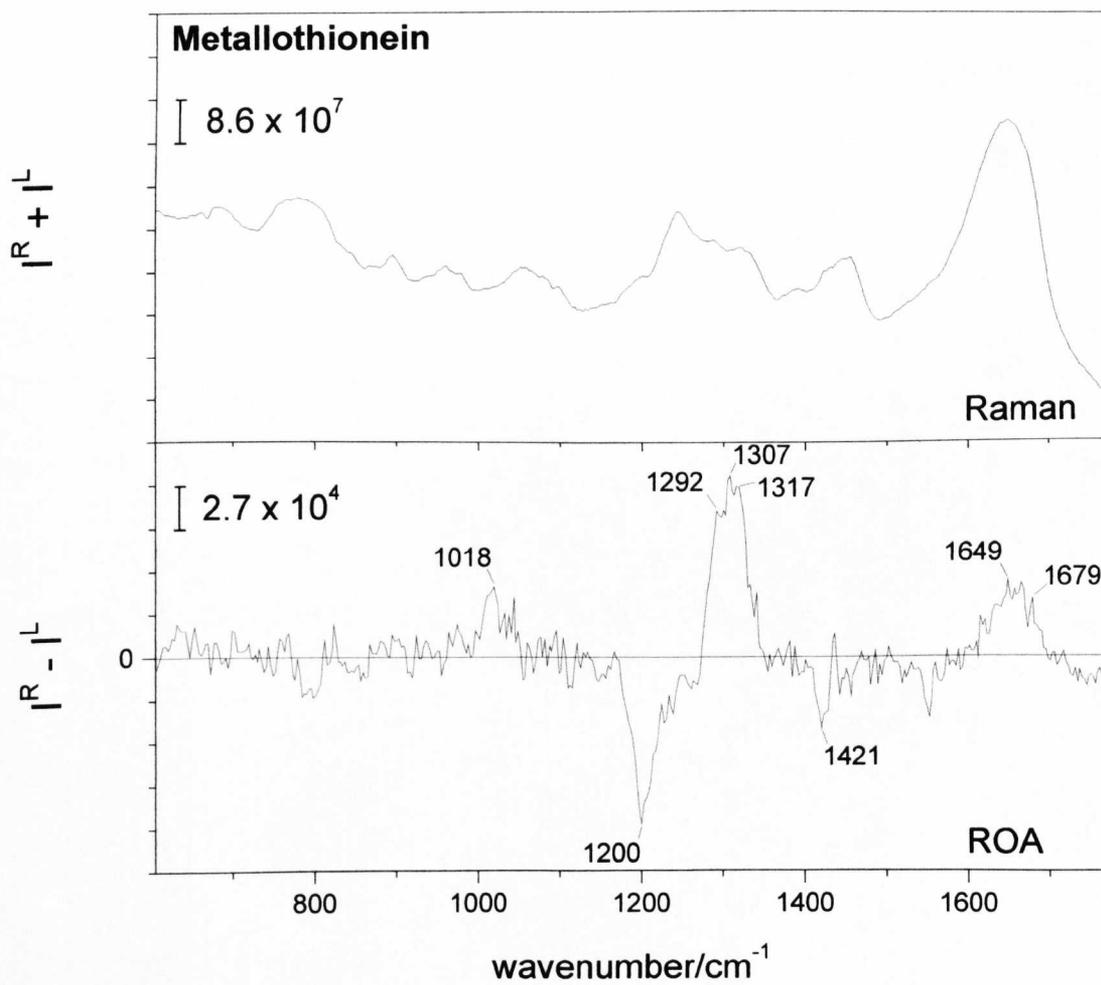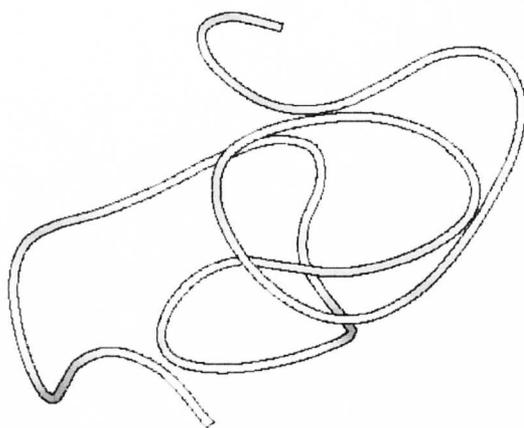
**Figure 6.3** Molscript diagram and the backscattered Raman and ROA spectra of Cd$_7$-metallothionein from rabbit liver in 0.05M Tris buffer at pH 8.0, ~67 mg/mL.

## Bovine orosomucoid

Bovine orosomucoid (or $\alpha_1$-acid glycoprotein) is a member of the lipocalin family of proteins. Lipocalins are transporters for small hydrophobic molecules, such as lipids, steroid hormones, and retinoids. (Kremer et al., 1988). Orosomucoid has a molecular weight of around ~45 kDa, and consists of 111 amino acids. At present there is no X-ray crystal structure or solution NMR structure available for this protein, however the SWISS-PROT classification system suggests that these proteins have a significant amount of $\beta$-strand, possibly similar to $\beta$-lactoglobulin which is also a member of the lipocalin family. However, bovine orosomucoid is known to contain a significant amount of bound carbohydrate (like invertase), perhaps as much as 40%.

The backscattered Raman and ROA spectra of bovine orosomucoid in aqueous solution are shown in Figure 6.4. In a clear difference to the ROA spectra of invertase and metallothionein, there is a strong sharp couplet in the amide I region, peaking at ~ 1678 cm$^{-1}$, strongly indicating the presence of a significant amount of $\beta$-sheet. Also, there is a sharp band at ~1243 cm$^{-1}$ which also suggests the presence of $\beta$- strand.

The amide III region is reasonably informative as well, suggesting a relatively high proportion of amino acid residues in long loops, with a positive peak at ~1293 cm$^{-1}$ tending to suggest that a number of residues exist in the general $\alpha$ region of the Ramachandran surface (as opposed to the presence of $\alpha$-helix itself), although there is also a sharp band of medium intensity at ~1338 cm$^{-1}$ that suggests the presence of a small amount of hydrated $\alpha$-helix. The view that the structure of orosomucoid may be similar to the well-known structures of other members of the lipocalin family (such as $\beta$-lactoglobulin) is supported by its homology with these proteins, similarities in disulfide bond arrangements, and its secondary structure profile, predicted from the amino acid sequence (Pervais et al., 1987). This being the case, orosomucoid may possibly contain up to ~12% $\alpha$-helix, and ~40% $\beta$-strand.
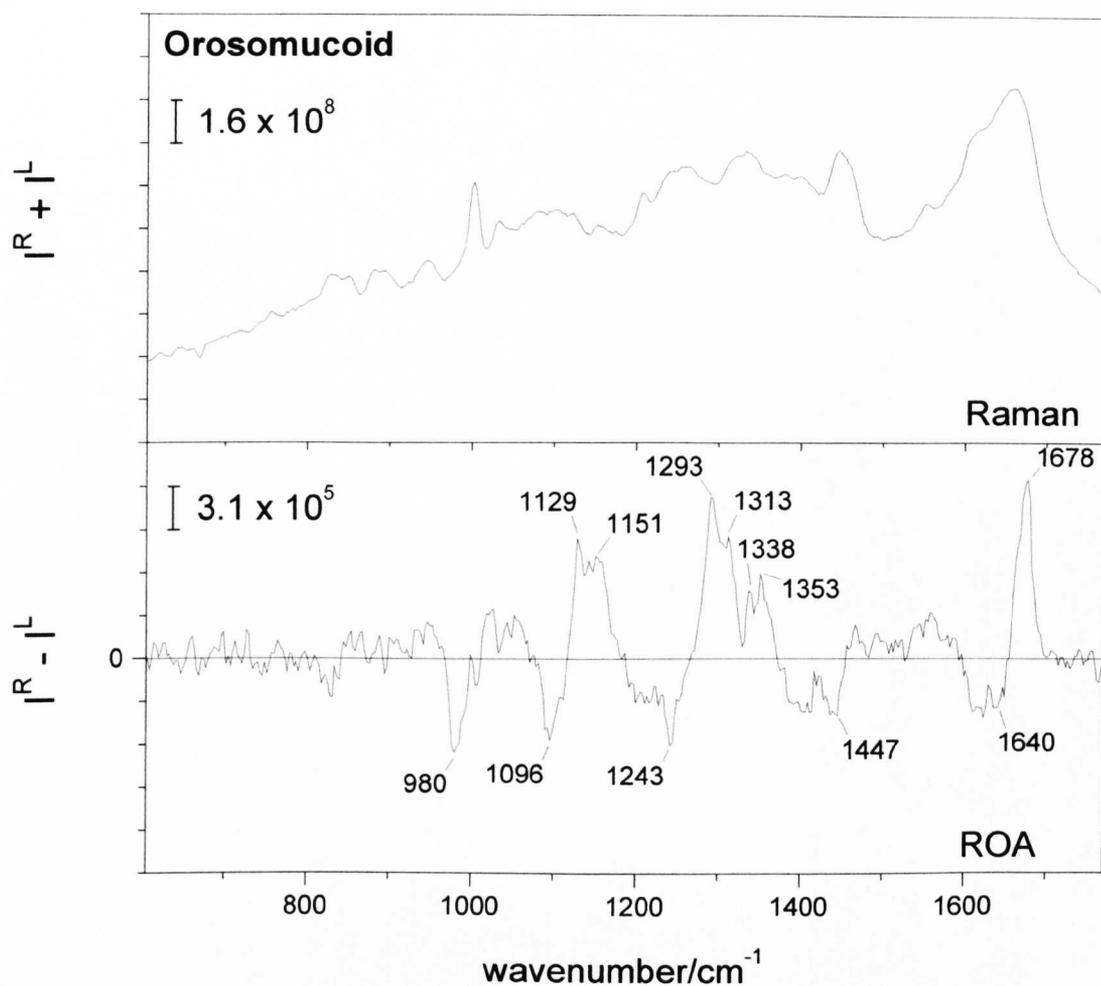
**Figure 6.4** The backscattered Raman and ROA spectra of bovine orosomucoid in 0.1 M acetate buffer with ~8mM KI, pH 5.4 and ~100 mg/mL.

The presence of strong ROA bands below ~1200 cm$^{-1}$ is thought to be due to the high carbohydrate content of this sample, much like the same region in the ROA spectrum of invertase discussed previously. The amide III region does not clearly resemble that of any groups of proteins with regular folds, and this may be due to a significantly higher amount of extended loop structure, and possibly PPII, although the peak at ~ 1313 cm$^{-1}$ may be a little low to support this idea. However, from ~1400 cm$^{-1}$ and upwards, a very close similarity can be seen between this protein and the well-known β-barrels as discussed in the previous chapters. The detail and sharpness of the structure in the ROA spectrum of bovine orosomucoid suggests that this protein exists in a well-defined native fold, yet its fold could well be classed as irregular, even when compared to the closely related lipocalin family members which are known to form β-barrels.

# Fish antifreeze protein (AFP) type III

Antifreeze protein type III belongs to a class of proteins that lower the freezing point of blood by absorbing ice and inhibiting its growth. Serum antifreeze proteins (or AFP's) are functionally similar yet structurally diverse proteins, with the type III proteins being rich in $\beta$-structure (Sonnichsen et al., 1996) The solution NMR structure of AFPIII is available (PDB code 3rdn) (Sonnichsen et al., 1996) and is illustrated as a Molscript diagram in Figure 6.6. According to the CATH classification system (Orengo et al., 1997), AFPIII has few secondary structures and does not belong to any of the regular fold classes discussed in previous chapters. The solution NMR structure suggests that there is a significant proportion of residues in extended loop and turn structure, and that the total secondary structure content is only ~30%.

The backscattered ROA spectrum of AFPIII in aqueous solution is shown in Figure 6.5 and immediately confirms the initial observations from the solution NMR structure. The most prominent feature of the spectrum is the positive structure in the amide III region, with a central sharp peak at ~1321 cm$^{-1}$, strongly indicating the presence of sections of PPII helix in the long loops within the protein. There are also small shoulders at ~1300 and ~1332 cm$^{-1}$ that may be reflecting the presence of a small amount of $\alpha$-helix. There is a sharp negative peak at ~1219 cm$^{-1}$ which indicates the presence of hydrated $\beta$-strand, although the couplet in the amide I region with a positive peak at ~ 1672 cm$^{-1}$ is fairly broad, only of medium strength and only has a small negative component which peaks at ~1638 cm$^{-1}$. This would suggest that the $\beta$-strand that is known to be present do not form very large or well-defined $\beta$-sheets. Also, there is no sign of negative structure beyond ~1350 cm$^{-1}$ indicating a lack of $\beta$-turn structure. It would seem in the case of this particular protein that the structure is an irregular fold which is dominated by the presence of sections of PPII helix as opposed to the $\beta$-rich structure suggested in the literature.
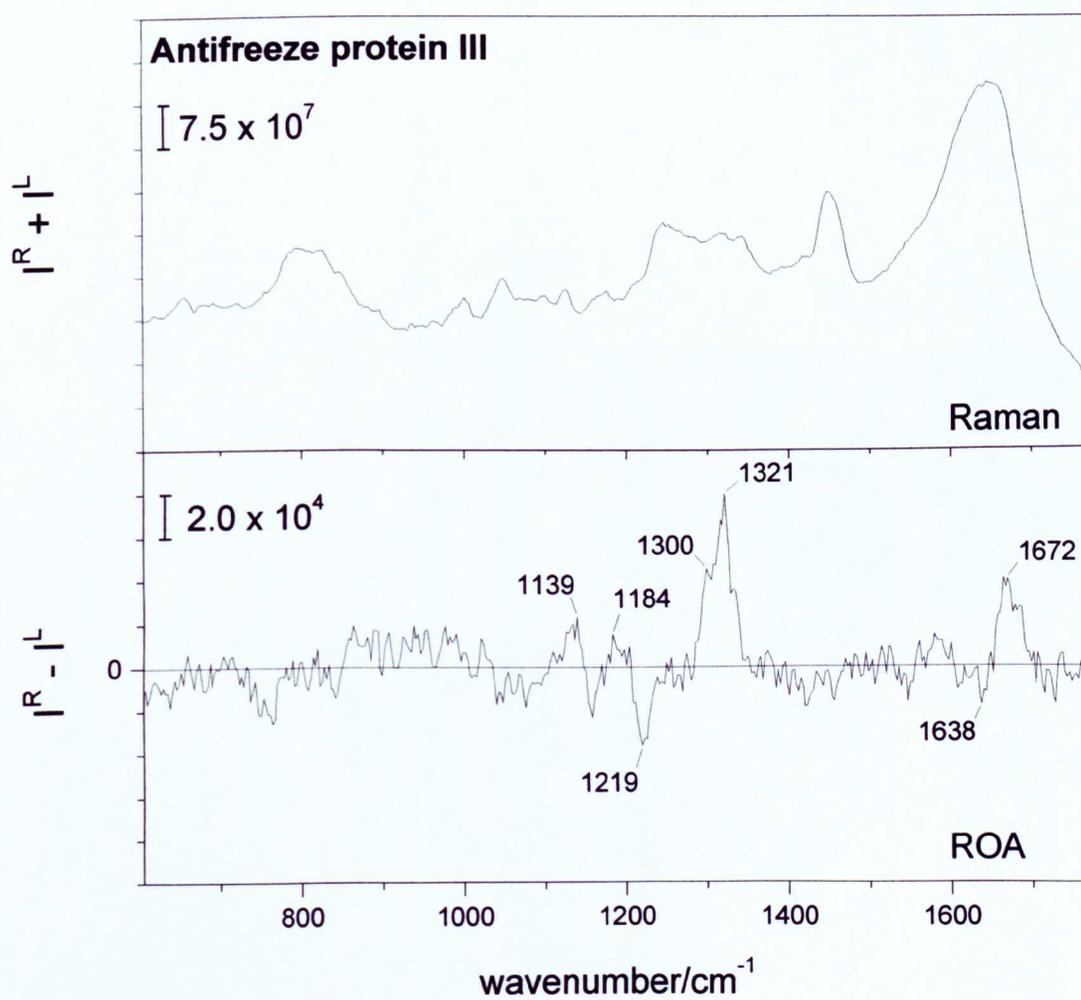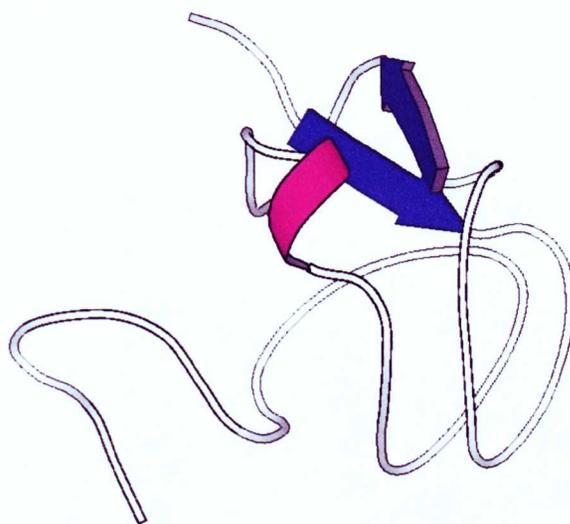
**Figure 6.5** Molscript diagram and the backscattered Raman and ROA spectra of antifreeze protein III in distilled water, pH 7.3 at ~50mg/mL.

## 6.4 Conclusion

The spectra presented in this chapter have afforded a new and useful insight into the structure of proteins with irregular folds, and allowed a distinction to be made between the nature of the disorder present in specific proteins. Perhaps the most significant conclusion is that the presence of a sharp positive peak at ~1320 $cm^{-1}$ which represents PPII helix, is seen in the model structures of disordered poly-L-lysine and poly-L-glutamic acid, and in the Bowman-Birk inhibitor, anti-freeze protein type III. However, this band is not seen in invertase, bovine orosomucoid or in metallothionein. In the first two cases, their irregular tertiary folds are stabilized by significant amounts of carbohydrate, and the structure of metallothionein is stabilized by the bound divalent metal ions.

# Chapter Seven

# Natively Unfolded Proteins

## 7.1 Introduction

Natively unfolded proteins have posed many difficulties to structural biologists and chemists. By their very nature, they are difficult to define in the usual terms as they contain little or no discernable secondary structure, and are generally not held together by any sort of co-operative fold assembly. The term 'natively unfolded' is used here to describe proteins whose native structures are non-regular, as opposed to the term 'unordered' that suggests a chemically induced state from a previously ordered state. A more commonly used term is random coil, the definition of which has become important in the discussion of unordered proteins. This term implies a randomness in which individual bonds between amino acid residues are free to rotate with some degree of hindrance (Poland et al., 1970). It is becoming increasingly apparent however that this extreme situation does not occur in most non-regular protein states. The fact that such proteins can have important biological functions is a key indicator that some sort of more static disorder is present.

In this chapter, a number of proteins that have been described as natively unfolded with random coil structures are presented. Among them, two are known to have a propensity for fibril formation and are involved in neurodegenerative disease. New insight into the structures of these proteins has been gained by using ROA to study their structures. It is immediately apparent that the ROA spectra of this type of protein are all very similar, echoing the similarity of proteins with the same fold types as demonstrated in previous chapters. In particular, a characteristic positive band at ~1320 cm$^{-1}$ is clearly seen in these spectra, thought to be arising from sections of PPII helix. The presence of a relatively large amount of PPII helix is believed to impart a rheomorphic (flowing shape) character to the structure of these proteins (Syme et al., 2002), which could be essential for their native function but which may also result in a propensity for pathological fibril formation in certain proteins (Spillantini et al., 1997; Goedert, 1999).

## 7.2 Experimental

Pure samples of β-casein and κ-casein (from bovine milk) were supplied as freeze dried powders by C. Holt of the Hannah Research Institute in Ayr. Solid samples of β- and γ-synuclein (human), as well as aqueous solutions of human tau 46 wild-type and P301L mutant were prepared and supplied by Dr. M. Goedert of the Medical Research Council, Cambridge. Hen phosvitin was obtained as lyophilized powder from the Sigma-Aldrich Company. In addition to these samples, it is necessary to include for the purposes of the discussion the ROA spectrum of α-synuclein (supplied by M. Goedert) although this spectrum was not obtained by the author. Similarly, the spectra of disordered poly-L-glutamic acid and poly-L-lysine are shown for reference only.

The casein samples were all analyzed in 50mM phosphate buffer at pH 7 at a concentration of 50 mg/mL. All samples were studied at room temperature (20°C), and in addition a study on β-casein was carried out with experiments at 4°C, 20°C and 50°C.

The synuclein samples were dissolved in 0.1M Tris buffer at an approximate concentration of ~150 mg/ml, although the ROA spectra suggest that the actual concentration is significantly lower. Large buffer salt bands were observed in the parent Raman spectra of all the synuclein samples, and artefacts in the ROA spectra were also observed as a result.

The samples of human tau 46 were supplied as 250 μl solutions in 0.1M Tris buffer at a concentration of ~ 3 mg/mL. The solutions were placed in Centricon vials (MW 10,000 cut-off membrane) and cold-centrifuged to an approximate concentration of ~30 mg/mL. During the ROA experiment, an oily layer began to develop at the bottom of the cell compartment, most probably due to aggregation in the beam. The pH was lowered to ~pH 4.3, and no aggregation occurred. The sample cell was periodically centrifuged to remove slight cloudiness due to aggregation after the sample had spent some time in the beam.

## 7.3  Results and discussion

### β- and κ-casein

The caseins make up almost 80% of proteins in bovine milk. The major components are $\alpha_{s1}$-, $\alpha_{s2}$-, β- and κ-casein and are present in proportions 0.37 : 0.09 : 0.41 : 0.13 respectively as colloidal calcium phosphate micelles (Holt, 1992; Rollema, 1992). The monomers have molecular weights of ~19-25 kDa, and are relatively unconstrained structures with few disulphide links which are inter- rather than intramolecular (Groves et al., 1992; Rasmussen et al., 1992). Early structural studies determined the caseins to be mostly structureless proteins with little or no secondary structure. Later, UVCD studies suggested that, although largely random coil, $\alpha_{s1}$- and β- casein may contain up to 20% α-helix and possibly some β-sheet (Creamer et al., 1981; Graham et al., 1984). A conventional Raman study (Byler & Susi, 1988) suggested that the two proteins were not identical, as differences in the fine structure of the two spectra were observed. UVCD and FTIR studies of κ-casein (Raap et al.,1983; Griffin et al., 1986; Ono et al., 1987) indicate the presence of ~10-20% α-helix, but a significant amount (~30-40%) of β-sheet as well. Sequence-based structure predictions suggest that the caseins are of the β-strand type, but that condensation into β-sheet is inhibited by certain features of the primary structure, allowing the caseins to retain an open and mobile rheomorphic conformation (Holt & Sawyer, 1993)

Figure 7.1. shows the backscattered Raman and ROA spectra of bovine β-casein (top pair) and κ-casein (bottom pair), measured at room temperature. The ROA spectra are very similar, showing that the basic structures of the proteins in aqueous solution are also very similar. The main feature in both ROA spectra is the central positive band at ~1318-1320 cm$^{-1}$ in the extended amide III region, similar to that observed in the model disordered homopolypeptides, and assigned to PPII helix (Smyth et al., 2000). The positive ROA bands at ~1290-1295 cm$^{-1}$ may originate in other types of loops and turns. The well-defined negative band at ~1245 cm$^{-1}$ in the ROA spectrum of κ-casein is assigned to β-strand, as opposed to the normal assignment of
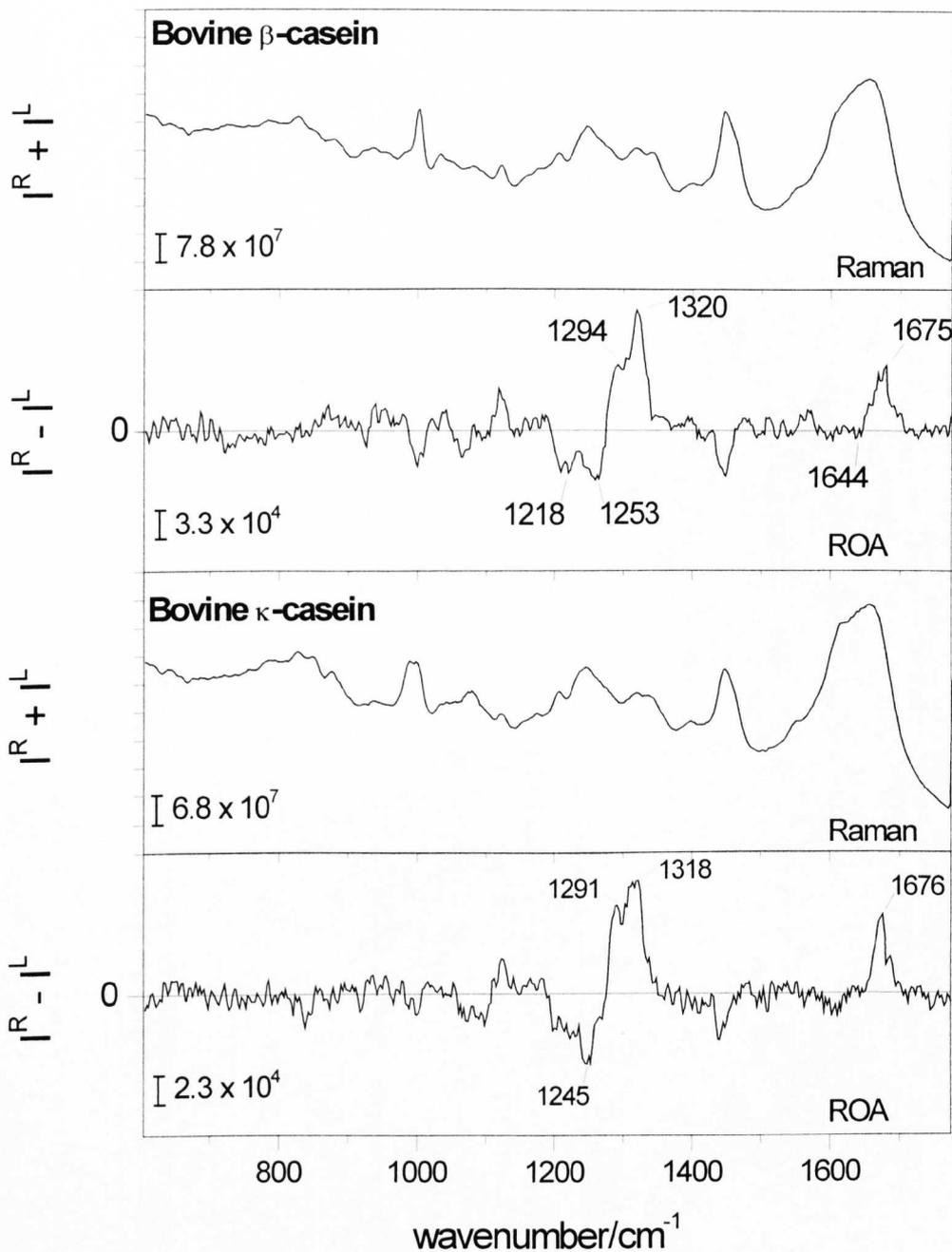
**Figure 7.1** The backscattered Raman and ROA spectra of bovine b-casein (top pair) and k-casein (bottom pair), measured at 20°C at pH 7.0 in 50mM phosphate buffer, 50 mg/mL.

β-sheet attributed to negative structure at ~1238-1253 cm$^{-1}$, due to the appearance of the amide I region. The negative intensity in β-casein in this region may have a similar origin. Both caseins show significant negative intensity at ~1220 cm$^{-1}$ which is believed to originate in a more hydrated form of β-strand (Smyth et al., 2000).

In the amide I region, both spectra show positive structure peaking at ~ 1675-6 cm$^{-1}$. Although at a high enough wavenumber to suggest some β-sheet content, the lack of a positive-negative couplet like that seen in the beta proteins suggests that the beta structure thought to be present is in the form of unassociated β-strands rather than β-sheet (Syme et al., 2002). In summary, these two ROA spectra suggest that the main conformational element in β- and κ-casein is PPII helix, possibly with some unassociated β-strand.

### α-, β- and γ-synuclein

The synucleins are a group of related brain proteins of unknown function from 127 to 140 amino acids in length (Jakes et al., 1994; Uéda et al., 1993; Ji et al., 1998). They are known to lack both cysteine and tryptophan residues and structural studies by UVCD and other techniques suggest that they adopt random coil structures (Weinreb et al., 1996; Serpell et al., 2000; Biere et al., 2000). α-synuclein is implicated in fibril formation in Parkinson's disease and other neurodegenerative illnesses (Goedert, 1999).

The backscattered Raman and ROA spectra of human α-, β- and γ-synuclein are shown in Figure 7.2. These samples contained a large concentration of buffer salts which gave rise to large bands in the parent Raman spectra (buffer salt bands are marked 'b'). These bands also gave rise to artefacts in the ROA spectra, and as a result, short segments of the data have been removed for clarity. Fortunately, these bands do not obscure any areas of great interest in these particular spectra, with the amide I and extended amide III regions being unaffected. The quality of these spectra is generally lower than that of the main body of proteins studied as part of this survey, and is due to a number of factors not least the unordered nature of the protein itself. So saying, these ROA spectra represent a new and unique insight into the structure of these difficult to obtain proteins and reveal a similarity to the other proteins in this chapter that suggests that PPII helix plays an important role in determining the properties of these proteins.
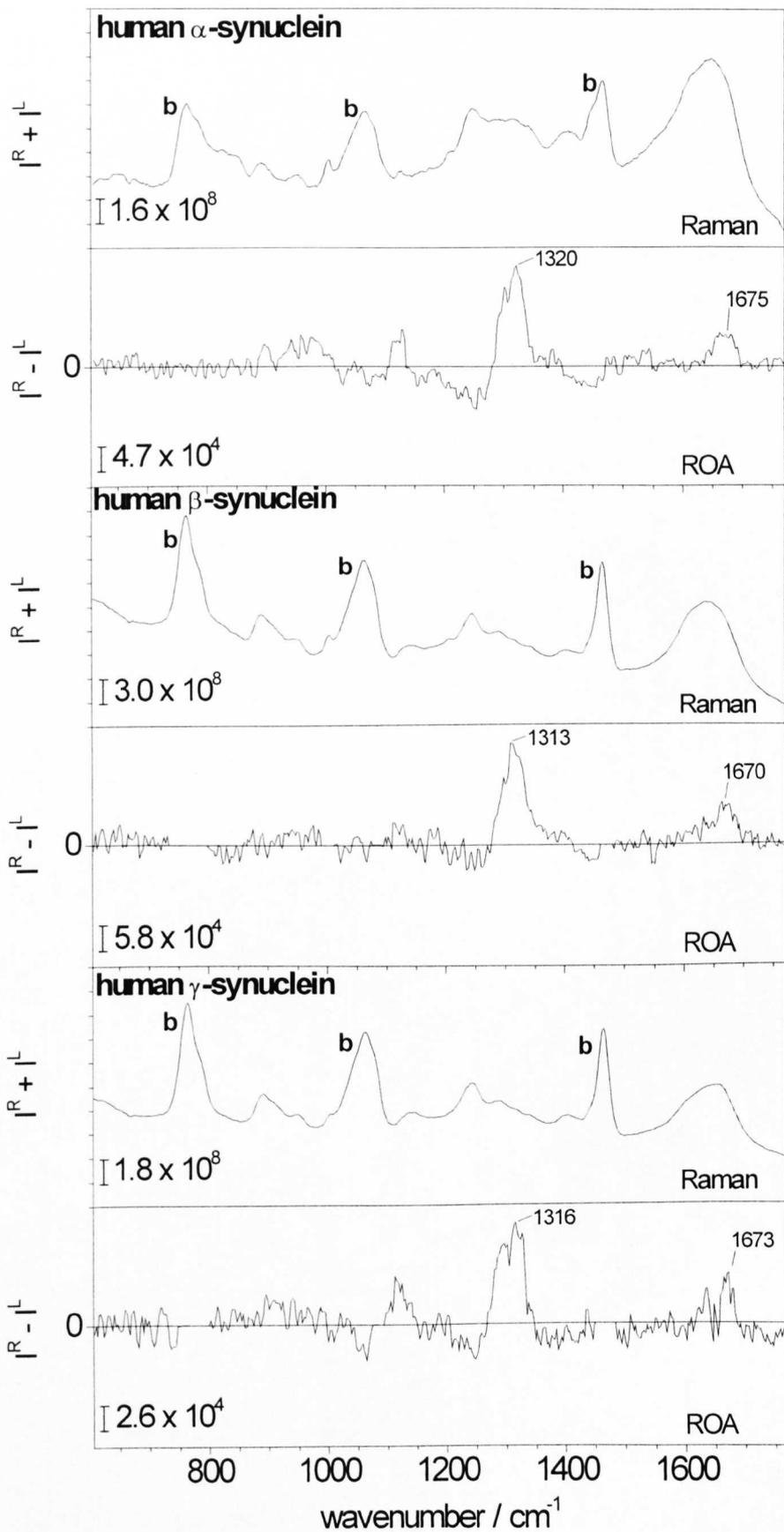
**Figure 7.2** The backscattered Raman and ROA spectra of α-synuclein (top pair), β-synuclein (middle pair) and γ-synuclein (bottom pair).

All three ROA spectra in Figure 7.2. are very similar to each other, each being dominated by a strong positive band centred at ~1313-1320 cm$^{-1}$ assigned to PPII helical structure. Also, in the amide I region, each has a lump of positive structure peaking at ~1670-1675 cm$^{-1}$ assigned to disordered/PPII structure. There is less definition than in the caseins, suggesting that little or no β-strand is present. There is no sign of negative structure at low wavenumber indicating that there is little or no secondary structure. These spectra suggest that the major conformational element in the synucleins, like the caseins, is PPII helix.

## Human tau 46 protein

Six isoforms of tau protein are expressed in the adult human brain, ranging from 352 to 441 amino acid residues in length (Goedert et al., 1989). Two isoforms, tau 37 and tau 46 were subjected to ROA studies, of which the wild-type and P301L mutants of tau 46 were most successful. Human tau protein has a random coil structure with very little secondary structure (Schweers et al., 1994; Goedert et al., 1999; Barghorn et al., 2000). Tau 46 P301L mutant causes frontotemporal dementia and Parkinsonism linked to chromosome 17.

The backscattered Raman and ROA spectra of wild-type and P301L mutant human tau 46 protein are shown in Figure 7.3. Both samples were run at lowered pH to alleviate problems with aggregation. Under such mild acidic conditions, it is unlikely that the conformation of the proteins have been altered significantly since the native proteins are already in an unfolded state. Both spectra show a strong positive band centered at ~1316-1318 cm$^{-1}$, indicating that PPII helix is likely the major conformational element in these proteins. Once again, the amide I region exhibits little similarity to that of the folded proteins of previous chapters, peaking at ~1670-1675 cm$^{-1}$, but with no clear strong positive peak, and no negative component whatsoever, highly suggestive of little secondary structure. However, there are negative peaks in the amide III region at ~1240-1266 cm$^{-1}$ that suggest there is possibly some β-strand present. The scarcity of additional bands throughout the rest of the spectra, particularly at lower wavenumber provides further evidence for the lack of regular secondary structure, and gives an overall impression of an open structure.
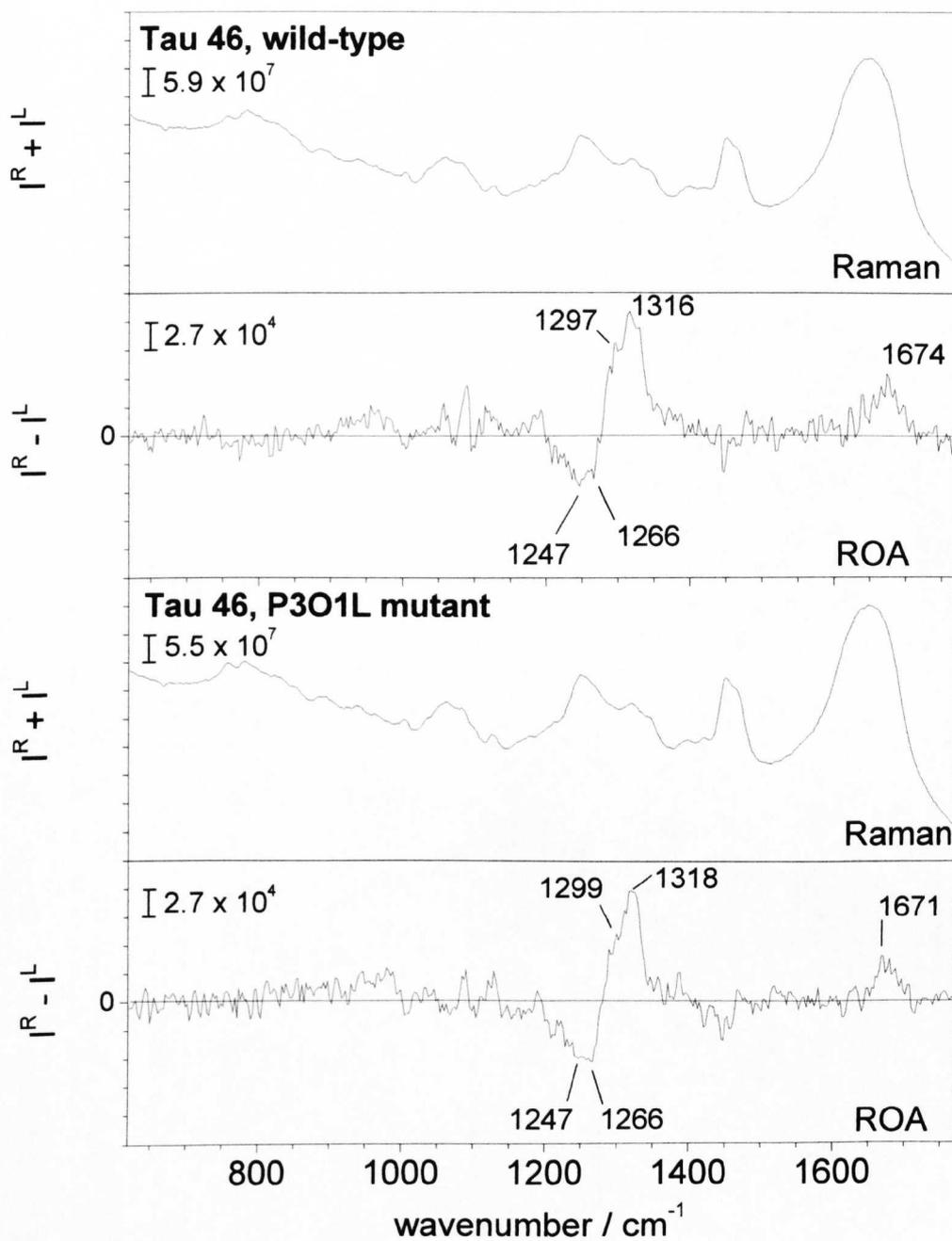
**Figure 7.3.** The backscattered Raman and ROA spectra of human tau 46 wild-type (top pair) and P301L mutant (bottom pair). Both spectra were measured at ~33 mg/ml in Tris buffer at ~ pH4.

**Hen phosvitin**

Phosvitin is a highly hydrophilic phosphoglycoprotein that constitutes the major protein in egg yolk and represents 1% of the mass. (Burley et al., 1989) Hen egg yolk phosvitin has a molecular weight of ~34 kDa with ~ 6.5% carbohydrate by mass. Phosvitin contains 216 amino acid residues of which some 123 are serines, most of which are phosphorylated. The crystal structure of phosvitin has not been solved, and various spectroscopic studies of the solution structure of phosvitin have suggested a number of possibilities. NMR (Vogel et al., 1983), conventional Raman (Prescott et al., 1986) and vibrational circular dichroism (VCD) (Yasui et al., 1990) all suggest that, at neutral pH, phosvitin has an irregular structure. UVCD and Fourier transform infrared (FTIR) (Losso et al., 1993) indicate the presence of some β-sheet in addition to irregular structure. At low pH however, it is thought that β-sheet is predominant.

Figure 7.4 shows the backscattered Raman and ROA spectra of hen egg yolk phosvitin. The most obvious observation to be made about the ROA spectrum of phosvitin is that there are no clear bands coming from sections of extended secondary structure. In particular, there are no significant features within the skeletal backbone stretch region and no couplet within the amide I region, negative at low wavenumber and positive at high, suggesting the absence of α-helix and β-sheet. There is a positive lump centered at ~1672 cm$^{-1}$ similar to that seen in the disordered polypeptides poly-L-glutamic acid and poly-L-lysine. This may arise from open turn conformations similar to β-strand or PPII structure.

The most prominent feature of the ROA spectrum of phosvitin is the large negative-positive couplet in the amide III region. The negative peak (possibly structured) centered at ~1270 cm$^{-1}$ may indicate the presence of some β-strand, but the positive peak centered at ~1312 cm$^{-1}$ is unusually narrow and suggests the presence of a smaller range of $\phi,\psi$ angles than those in the reduced proteins and maybe a consequence of the large number of phosphorylated serine residues. Previous investigations of phosvitin by VCD suggest a similar conformation to disordered poly-L-glutamic acid and poly-L-lysine, in particular PPII structure. However, the ROA band centered at ~1312 cm$^{-1}$ seems a bit low to be assigned to PPII structure. At

low pH, where phosvitin is believed to adopt a β-sheet formation, the phosvitin solution turns into a gel and as a result it has not been possible to acquire an ROA spectrum at low pH.
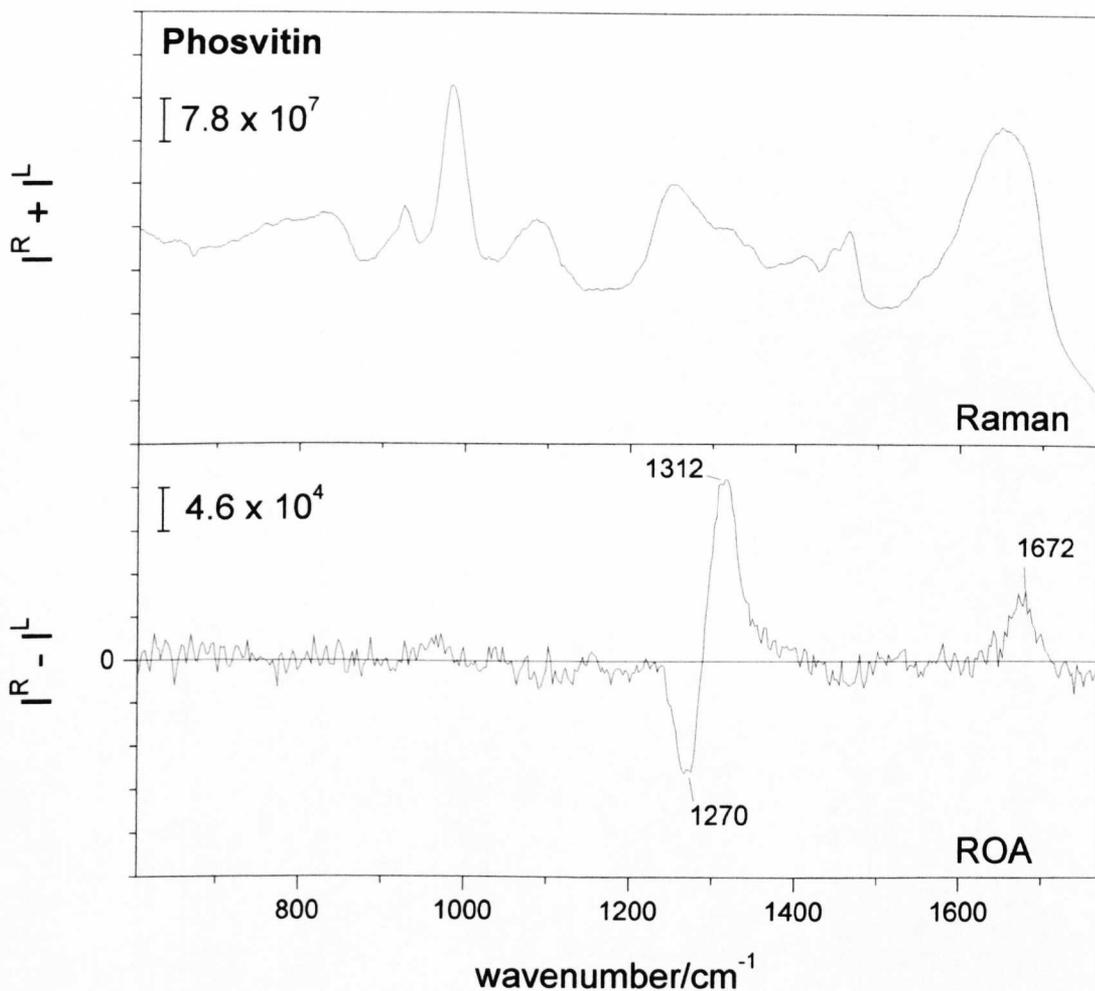


**Figure 7.4** The backscattered Raman and ROA spectra of hen egg yolk phosvitin in 0.1M acetate buffer at pH 5.4, at 100 mg/mL.

## 7.4 Conclusion

The protein spectra presented in this chapter are all very similar in their lack of detail pertaining to the assignment of secondary structure. These spectra also show strong evidence for the presence of PPII helix with a strong positive ROA band at ~1320 cm$^{-1}$ similar to that seen in the spectra of the disordered homopolypeptides poly-L-glutamic acid and poly-L-lysine upon which the PPII helix ROA band assignment is based. The rheomorphic (or flowing shape) character imparted onto

tau, casein and synuclein proteins by large amounts of extended, flexible PPII helix may be important for the function of these proteins and as such suggests that ROA may be a valuable tool in studying their role in neurodegenerative illnesses such as Alzheimer's disease and the prion diseases vCJD and BSE.

The characteristic band at ~1320 cm$^{-1}$, the positive 'lump' (but no clear peak) in the amide I region, the absence of a negative band in the amide I region, and general lack of detail especially at lower wavenumber, all serve to make the ROA spectra of these proteins easily distinguishable from any other groups of proteins seen in this thesis. As will be demonstrated by the results presented in the following chapter, these features (or lack of them) are useful in discriminating between proteins that are mainly ordered/structured and proteins that are largely (or completely) disordered.

The application of ROA to the study of this type of protein in particular highlights the usefulness of the technique for looking at proteins where very little is known about the structure of the protein from alternative methods. The fact that small amounts of secondary structure that are believed to or known to be present still influence the shape of the ROA spectrum suggests that ROA is extremely sensitive even when the protein is mostly disordered or does not exist in a regular folded state.

# Chapter Eight
# Principal Component Analysis of Protein ROA Spectra

## 8.1 Introduction

Until recently, the analysis of Raman optical activity spectra has been conducted by visual examination. The relationship between ROA spectral features and protein structural properties has been established on this empirical basis with much success, but has remained difficult to put into any sort of quantitative terms. For this reason, a more rigorous method of analysis has been sought and a number of techniques have been considered. The most promising of these techniques is principal component analysis (PCA), and an algorithm designed for the analysis of protein ROA spectra is currently being developed in collaboration with Dr. K. Nielsen of the Danish Technical University.

PCA is a multi-variant statistical approach of factor analysis which allows a data set, such as the set of protein ROA spectra presented in this thesis, to be analyzed on the basis of the variations within the original data from the mean. PCA determines a set of abstract sub-spectra that, when recombined in a manner specific to each particular protein, recreates the shape of the ROA spectrum of the protein. The major advantage of this approach is the original data need not be processed or influenced in any way by the analyst prior to analysis. PCA has previously been applied to protein structure determination by both VCD (Pancoska et al., 1991) and UVCD (Venyaminov & Yang, 1996) which suggested that it may be useful in the analysis of protein ROA data. The PCA method described here is based on one used for the analysis of conventional Raman spectra of parchment (Nielsen et al., 1999).

The raw ROA spectral data for the whole data set of circa 56 proteins can be expressed as a matrix, which may be transformed into a set of orthonormal sub-spectra (or basis functions). These basis functions are calculated as eigenvectors, and their importance in terms of the contribution to the total variation within the data set is weighted by the eigenvalues associated with each eigenvector. Since the basis functions are of decreasing importance to the overall shape of the reconstructed

spectrum, the lowest order basis functions (indicated by the lowest eigenvalues) are discarded, as their contribution recreates experimental noise. Once the number of significant eigenvalues has been established, the coefficients $c_{ij}$ are calculated such that each protein ROA spectrum may be represented as an algebraic sum of the coefficients and their corresponding basis functions.

## 8.2 Basic theory

By arranging the reference spectra (the experimental data) into a matrix with rows corresponding to wavelengths ($\lambda$) and columns ($i$) for each individual protein in the data set, the number of independent components in our spectral series is equal to the rank $p$ of the matrix which can be found by using PCA (Pancoska, 2000). Spectral intensities are considered as variance from the mean value. The treatment of the data by the PCA algorithm is based upon matrix algebra where the matrix of experimental data is represented as $[\Theta_i(\lambda)]$. The decomposition of any ROA spectrum $[\Theta_i(\lambda)]$ ($i$=1,....,N), where N is the total number of spectra in the data set, from the reference set into the linear combination of $p$ common independent subspectra $\phi_j(\lambda)$ ($j$=1,...., $p$) is:

$$\Theta_i(\lambda) = A_i \sum_{j=1}^{p} c_{ij} \phi_j(\lambda) \tag{8.1}$$

where

$$A_i = \sqrt{\int_{\lambda_1}^{\lambda_2} \Theta_i^2(\lambda) d\lambda} \tag{8.2}$$

is a normalization constant that ensures all spectra contribute equally the overall variance space. The correlation matrix $[R]$ of the experimental data required to solve the eigenvalue problem is calculated as

$$[R] = [\omega_i(\lambda)]^T [\omega_i(\lambda)] \tag{8.3}$$

where

$$\omega_i(\lambda) = \frac{1}{A_i}\Theta_i(\lambda) = \sum_{j=1}^{p} c_{ij}\phi_j(\lambda) \qquad (8.4)$$

and $T$ is the transpose operator. The elements $r_{ij}$ of the correlation matrix $[R]$ are the numerically calculated values of the overlap integrals for all spectral pairs in the reference set:

$$r_{ij} = \int_{\lambda_1}^{\lambda_2} \omega_i(\lambda)\omega_j(\lambda)d\lambda \qquad (8.5)$$

The principal components of $[\Theta_i(\lambda)]$ are found by diagonalization of matrix $[R]$ that yields eigenvector matrix $[q]$ and diagonal eigenvalue matrix $\Lambda_{ij}\delta_{ij}$ such that:

$$[q]^{-1}[R][q] = [\Lambda_{ij}\delta_{ij}] \qquad (8.6)$$

Since $[R]$ is a real symmetric matrix, $[q]^{-1} = [q]^{T}$. The solution of equation 8.4 can be found by multiplying the normalized experimental data matrix by unit matrix $[q][q]^{T}$:

$$[\omega_i(\lambda)] = [\omega_i(\lambda)][q][q]^{T} \qquad (8.7)$$

From equations 8.4 and 8.7 we then get the basis functions $[\phi_j(\lambda)]$ and expansion coefficients $[c_{ij}]$:

$$[\phi_j(\lambda)] = [\omega_i(\lambda)][q] \qquad (8.8a)$$

$$[c_{ij}] = [q]^{T} \qquad (8.8b)$$

These equations transform the set of measured protein ROA spectra $[\Theta_i(\lambda)]$ into the diagonal matrix of eigenvalues $[\Lambda_{ij}\delta_{ij}]$. $p$ is defined as the minimal number of subspectra needed to reconstruct the experimental spectra using equation 8.1. For the purposes of this work, the criteria by which $p$ is selected is such that spectra are

reconstructed to a preset level of accuracy of ~97%. In practice, the number of basis function required to achieve this is ~20 from an original reference set of 56 protein ROA spectra.

## Diagrammatical representation

Figure 8.1 is a simple visual depiction of the processes involved in generating a set of abstract functions (the basis functions) and expansion coefficients from the experimental data set containing $N$ spectra. As per equation 8.3, the matrix of protein ROA spectral data (depicted here as D) is multiplied by its transpose to give the correlation matrix [$R$]. Solving the eigenvalue problem (equations 8.5 to 8.7) yields the matrices detailed in equations 8.8a and 8.8b corresponding to the set of basis functions (B) and corresponding expansion coefficients (C), where $p$ may be selected



**Solving the eigenvalue problem**
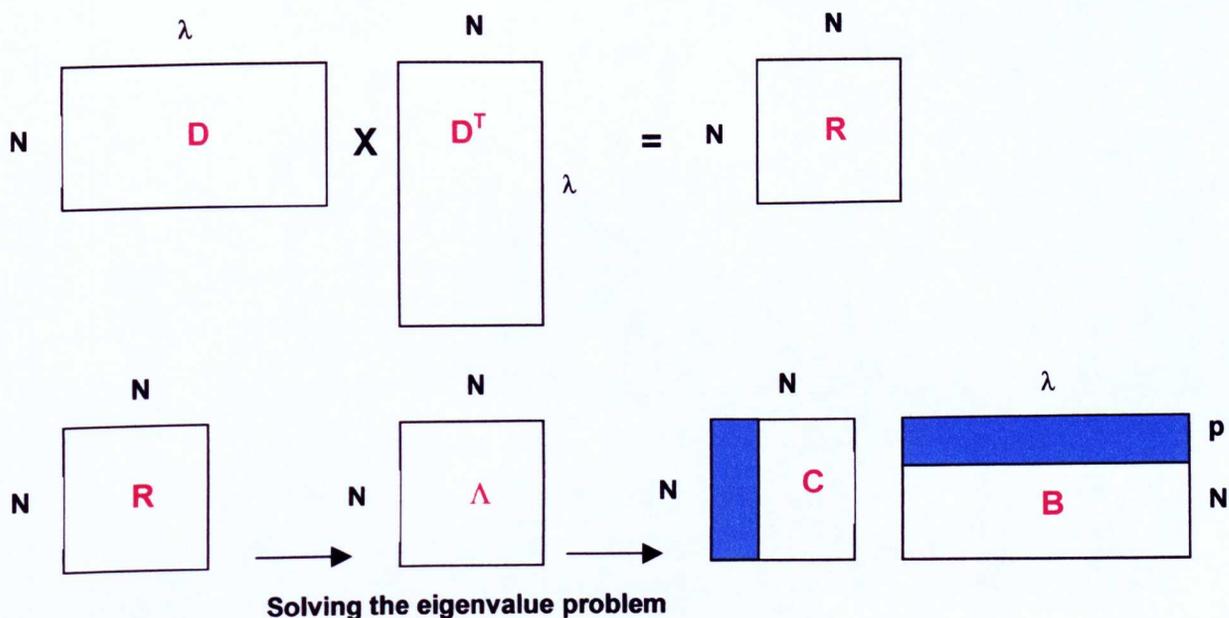
**Figure 8.1** Visual description of the matrix treatment of ROA data illustrating how the calculated basis functions and expansion coefficients relate to the real spectral data.

to achieve the desired level of accuracy. At this point is it useful to realize that the matrix B is of the same dimensions as [$\Theta_i(\lambda)$] representing the original spectra ($N \times \lambda$) and that the basis functions themselves appear visually similar to ROA data. Thus far,

point, the basis functions are merely abstract representations of the original data and do not possess any real physical significance.

To determine whether or not the raw PCA treatment of the data has any correlation to physical properties of the analytes, it is necessary to consider the results of the PCA by analysis of the coefficients yielded for each protein. From matrices B and C, each protein will have $n$ associated coefficients, one each per $n$ basis functions required. The basis functions are ordered in terms of relative significance to the overall variance. Therefore, if the expansion coefficients (for all analytes) associated with one basis function are plotted against another, it is possible to visualize whether there is any relationship between the basis functions and some properties of the analytes.

## 8.3 Results and discussion

### 8.3.1 Raw PCA

The expansion coefficients pertaining to the two most significant basis functions were plotted for all 56 polypeptide, protein and virus samples in the data set and the result in shown in Figure 8.2A. Each point on the coefficient plot represents one of the 56 ROA spectra in the data set. In addition, each point is coloured to indicate the structural class of each protein as defined from the literature. This plot clearly shows clustering of points with respect to structural class. The first two basis functions between them should demonstrate the greatest degree of variation between individual spectra. In other words, this particular plot would be expected to reveal the best separation between clusters of proteins of different structural class. In Figure 8.2B, the points are numbered to indicate which protein each point represents in accordance with the names of the proteins shown in Table 8.1.

The first four basis functions are illustrated in Figure 8.3. Their appearance is very similar to that of typical protein ROA spectra, but the detail contained within each is difficult to interpret in terms of protein structure.
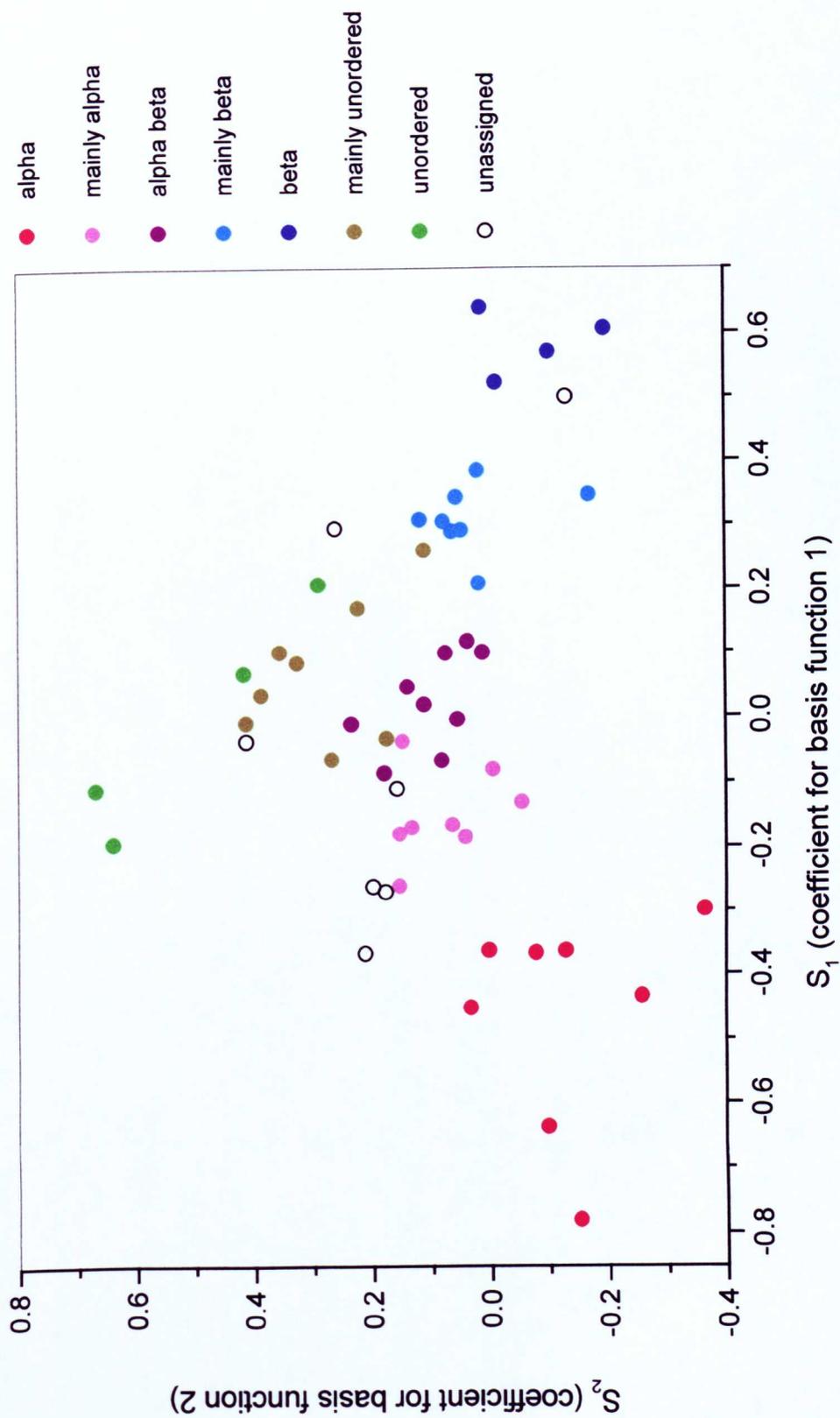
**Figure 8.2A** Coefficient plot for basis functions 1 and 2 from the raw PCA treatment of the ROA spectra of 56 polypeptide, protein and virus ROA spectra. Each protein has been classified into a structural class as determined from the literature.
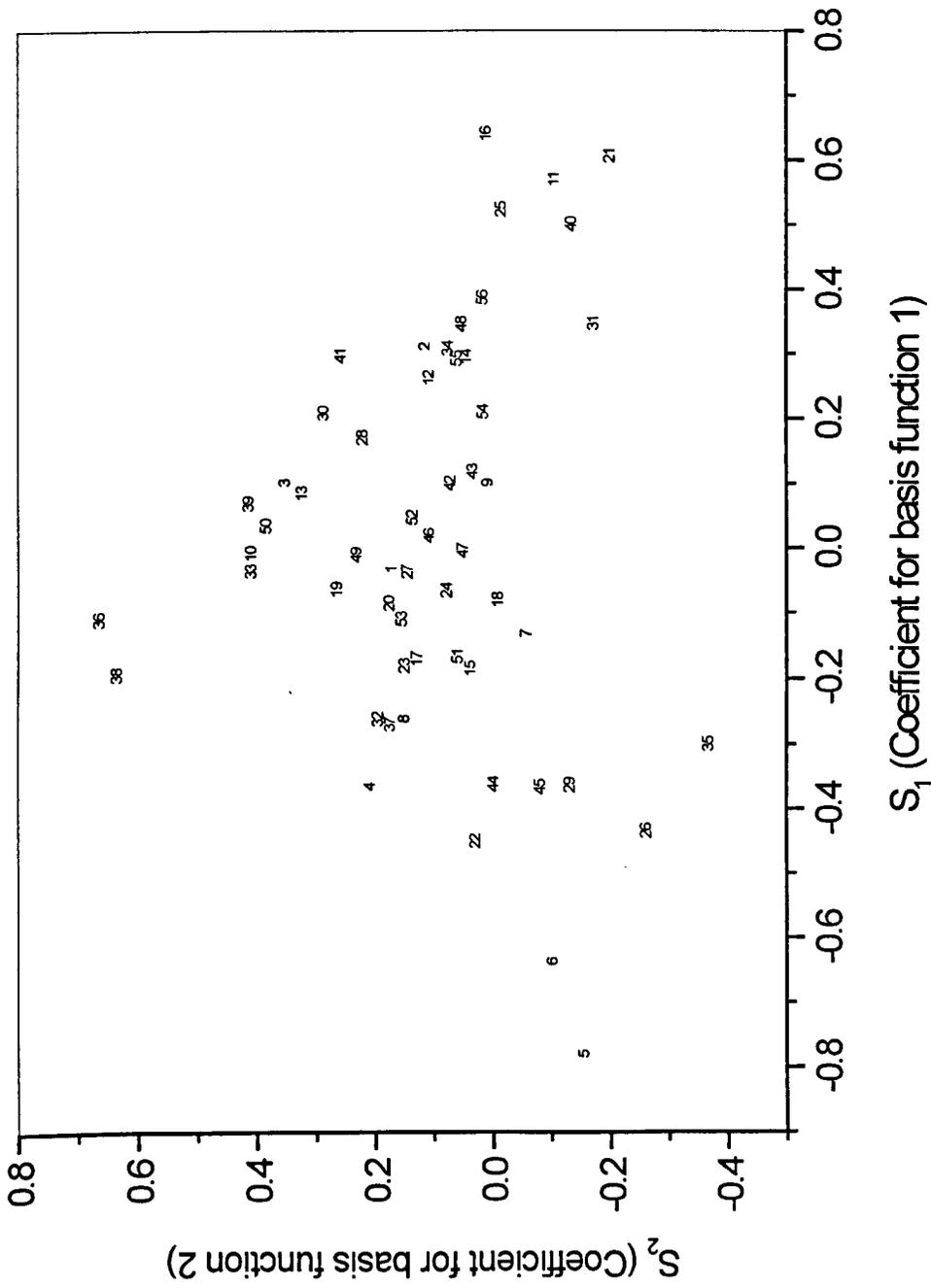
**Figure 8.2B** Plot showing the positions of each of the 56 proteins and polypeptides in Figure 8.2A. Table 8.1 lists the names of the samples

## Table 8.1  PCA Plot Reference Key

| No. | Protein Name | Abbrev. Code |
| --- | --- | --- |
| 1 | casein, alpha- | ACAS |
| 2 | chymotrypsin, alpha- | ACH |
| 3 | antifreeze protein 3 | AFP |
| 4 | gliadin, alpha- | AGLI |
| 5 | poly-L-glutamic acid, alpha-helical | AHPG |
| 6 | poly-L-lysine, alpha-helical | AHPL |
| 7 | lactalbumin, alpha- | ALAC |
| 8 | aldolase | ALD |
| 9 | amylase, alpha- | AMY |
| 10 | synuclein, alpha- | ASYN |
| 11 | avidin | AVID |
| 12 | bowman-birk inhibitor | BBI |
| 13 | casein, beta- | BCAS |
| 14 | lactoglobulin, beta- | BLB |
| 15 | creatine kinase | CK |
| 16 | concanavalin A | CONA |
| 17 | lysozyme, equine | EQL |
| 18 | lysozyme, hen | HEWL |
| 19 | lysozyme, human (prefibrillar) | HLPF |
| 20 | ovomucoid, hen | HOW |
| 21 | SAP, human | HSAP |
| 22 | serum albumin, human | HSA |
| 23 | lysozyme, human | HULY |
| 24 | ovalbumin, hen | HVA |
| 25 | immunoglobulin, human | IGG |
| 26 | IKE coat protein | IKEJ |
| 27 | insulin (monomeric) | INS |
| 28 | casein, kappa- | KCAS |
| 29 | M13 coat protein | M13 |
| 30 | metallothionein | MET |
| 31 | MS2 coat protein | MS2 |
| 32 | NMV coat protein | NMV |
| 33 | gliadin, omega- | OGLI |
| 34 | pepsin | PEP |
| 35 | PF1 coat protein | PF1 |
| 36 | phosvitin | PHO |
| 37 | potex virus coat protein | PVX |
| 38 | poly-L-glutamic acid, random coil | RCPG |
| 39 | poly-L-lysine, random coil | RCPL |
| 40 | reduced lysozyme | RELY |
| 41 | reduced ribonuclease A | RERA |
| 42 | ribonuclease A | RIB |
| 43 | ribonuclease B | RIC |
| 44 | S100 protein | S100 |
| 45 | S100B protein | S100B |
| 46 | S23 prion protein | S23 |
| 47 | S90 prion protein | S90 |
| 48 | satellite TMV coat protein | STMV |
| 49 | subtilisin carlsberg | SUB |
| 50 | tau46, human wild-type | TAU46 |
| 51 | TMV coat protein | TMV |
| 52 | ovomucoid, turkey | TOW |
| 53 | TRV coat protein | TRV |
| 54 | trypsin | TRY |
| 55 | trypsinogen | TSG |
| 56 | ubiquitin | UBI |

However, the basis functions do give a useful insight into the main factors within the data set as a whole, which most directly influence the overall shape of protein ROA spectra. It is not surprizing that the main bands in the most significant basis functions do correlate well with well-established ROA bands that have been assigned to types of structure. For example, basis function 1 has a prominent negative band centered at ~1340 cm$^{-1}$. The positive protein ROA band at ~1340 cm$^{-1}$ is associated with hydrophilic $\alpha$-helix, and we see from the plot shown in Figure 8.2. that highly helical proteins have large negative expansion coefficients for this basis function, the effect of which is to generate a large positive contribution at ~1340 cm$^{-1}$ in the reconstructed spectra of these proteins. There are several other examples of this type of correlation as well, such as the positive band at ~1320 cm$^{-1}$ in basis function 2, which can be associated with PPII helix. The protein ROA spectra of unordered proteins that contain large amounts of PPII helix are characterized by a prominent
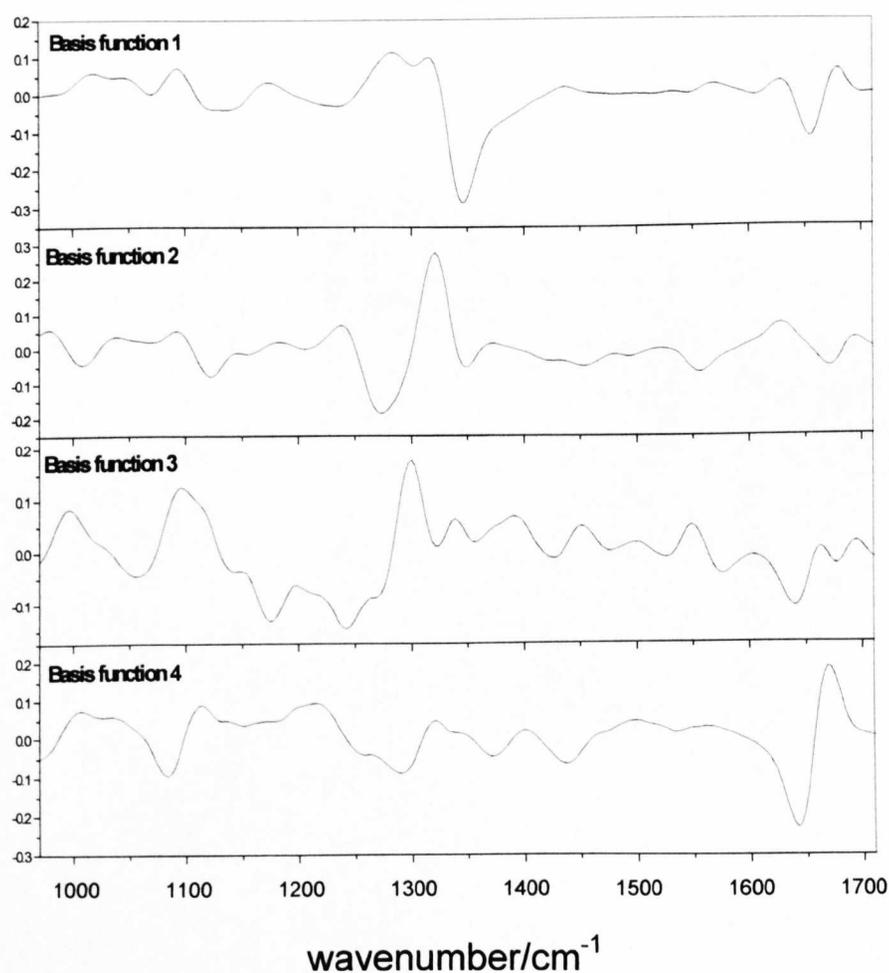


**Figure 8.3.** The four most significant basis functions as calculated by the raw PCA algorithm

positive band at ~1320 cm$^{-1}$, hence these proteins have large positive coefficients. This is also demonstrated in the plot in Figure 8.2., as the coefficients for basis function 2 are plotted along the Y-axis.

With reference to the first two basis functions, and considering the plot shown in Figure 8.2., it is tempting to say that the proteins are separated on the basis of helix content along the X-axis, and order/disorder along the Y-axis. This is generally correct, but it is important to stress at this point that the proteins are really mainly separating on the basis of positivity at ~1340 cm$^{-1}$ along the X-axis, and positivity at ~1320 cm$^{-1}$ along the Y-axis. In summary, it is the appearance of the clustering of points that is the starting point for the discussion of the results from the raw PCA algorithm. For the first time, it is possible to say that there exists some mathematical basis to describe the difference between the appearance of the ROA spectra of one structural class of protein from another. With the ability to separate clusters of points on the basis of general structural class from the raw PCA calculations, the next step is to be able to incorporate information from lower order plots simultaneously with other plots to provide more specific information, such as fold type and quantitative secondary structure content.

## 8.3.2 Trained PCA - orthogonal transformation of basis functions

There are two main ways to extract more specific information from the results of the PCA calculations. The first, as alluded to earlier, is to develop a way of analyzing multiple coefficient plots simultaneously (e.g. by non-linear mapping). The second is to change the way in which the PCA calculations are carried out, by introduction constraints into the algorithm and rotating the matrices in order to 'train' the PCA algorithm to generate plots of coefficients that optimize certain pre-defined characteristics of the analytes. This work has been carried out with the assistance of Dr. K. Nielsen and is currently work in progress, although some initial details and results will be described briefly here.

A main goal in the development of a new method for the analysis of ROA spectra has been to keep the analysis as objective as possible. The raw data calculated by the PCA algorithm discussed previously is probably the most objective analysis
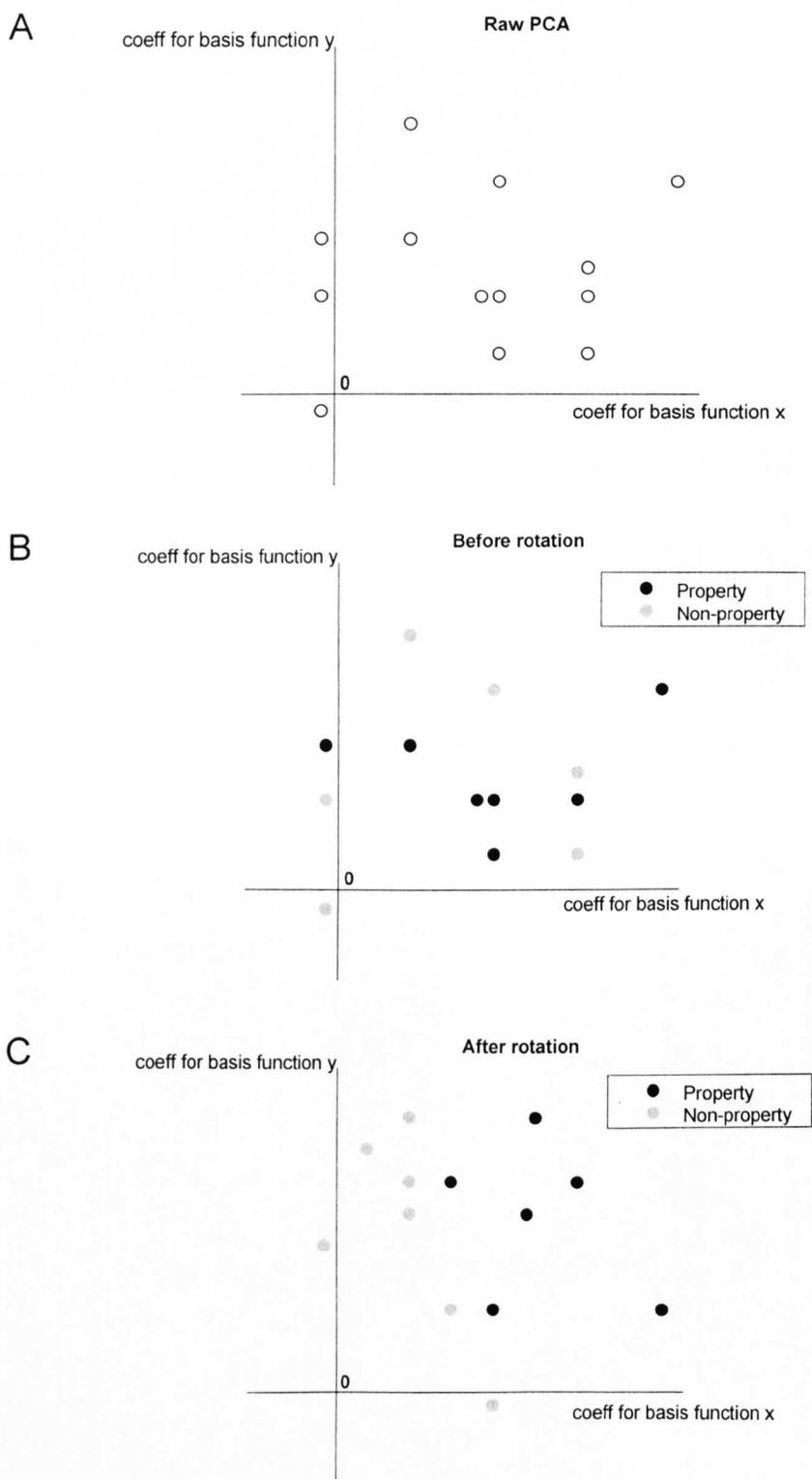
**Figure 8.4.** Diagrammatic representation of matrix rotation after orthogonal transformation. A – Raw PCA shows a random distribution of points. B – A property, such as helix content, is defined, and each point (protein) is classified as either having that property, or not having that property. C – The data can be presented on a new basis such that points marked as containing the desired property tend towards 1 on a particular axis (in this case, the X-axis) and towards zero for non-property proteins.

possible in that the data require no pre-treatment, and the results are generated without constraints. However, the discussion of the coefficient plots still depends upon a level of pre-knowledge with regard to the structures of the proteins involved. Orthogonal transformation of the matrices involves incorporating this pre-knowledge into the calculation of the basis functions by selecting properties that individuals within the data set either have (1) or do not have (0). The rotation of matrices is then carried out by the PCA algorithm, such that successive cycles of the program serve to maximize the separation of points marked 1 and points marked 0. The diagrams in Figure 8.4 are used as a visual guide to the effectiveness of this approach.

No conclusive results have been obtained by this method at the time of writing. However, the main properties of interest have been, to some extent, established and attempts to refine the discrimination have centered upon these properties. These properties include structural class, tertiary fold, relative secondary structure content, hydrophobicity, and amino acid composition. As more and more samples are subjected to analysis by ROA spectroscopy, the number of different properties that are having an influence on the shape of protein ROA spectra increases. It is the main objective of this analysis to be able to determine how each physical property of each individual sample contributes to the observed band patterns in ROA spectra.

### 8.3.3 Applications of raw PCA

From discussions of the appearance of protein ROA spectra such as those presented in previous chapters, and also from the initial results of the raw PCA treatment of these spectra, it is increasingly apparent that it is possible to gain very useful information simply from the shape of the ROA spectrum of a protein. The most obvious application of this new technique is that of fold prediction. At this early stage, it is possibly more realistic to attempt structural class determination. By incorporating a new protein ROA spectrum to the data set and using PCA to establish its relationship to the rest of the data, an initial indication as to the structural class, and to a certain extent an estimate of secondary structure may be made.

An example of this has already been published (Blanch et al., 2002) and applies to viruses, specifically potato virus X (PVX), narcissus mosaic virus (NMV) and tobacco mosaic virus (TMV). In addition, satellite tobacco mosaic virus (STMV) was also the subject of a recent ROA study (Blanch et al., 2002). These samples are of great interest to structural virologists, and little is known about the structures of PVX and NMV. The coat proteins of TMV are known to be mostly α-helical in the form of a four-helix bundle whereas STMV is known to be mostly β-sheet in the form of a jelly roll. Figure 8.5 is a modified version of Figure 8.2. and highlights the positions of these four viruses relative to the rest of the data set.
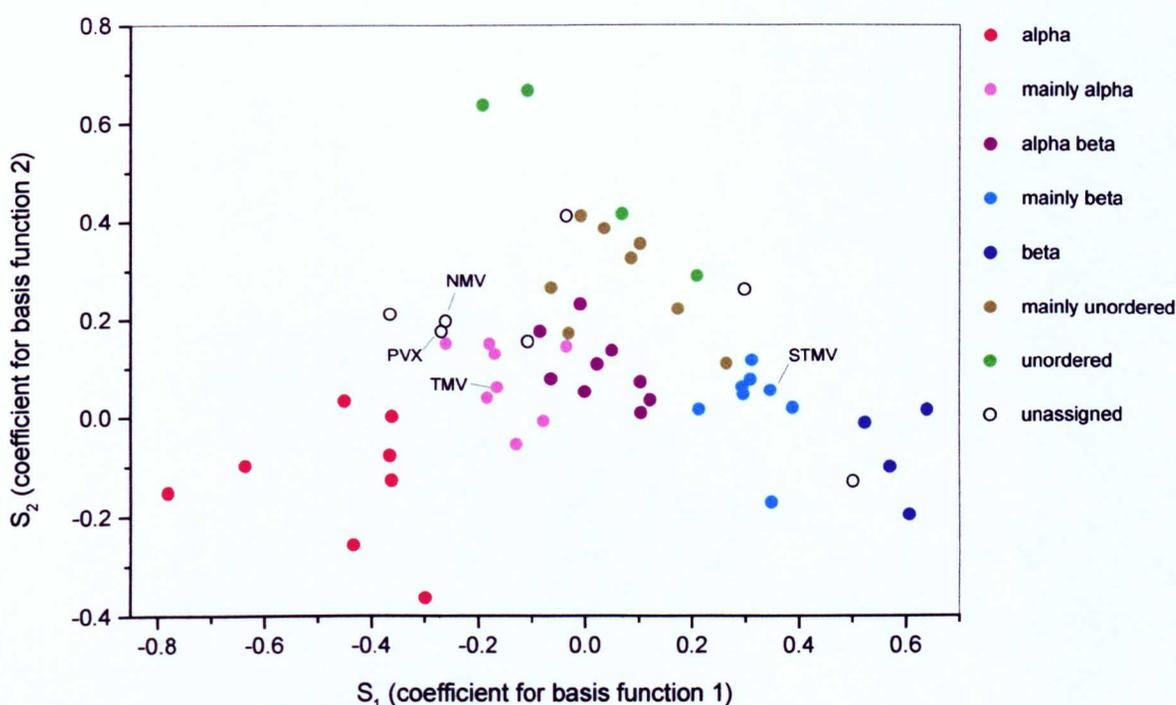


**Figure 8.5** Raw PCA plot showing the positions of four virus samples. TMV is known to be mainly alpha and appears in the mainly alpha cluster. STMV is mainly beta and appears in the mainly beta cluster. The previously unassigned PVX and NMV are thought to be mainly alpha (Sawyer et al., 1987; Blanch et al., 2002)

As this specific example demonstrates, raw PCA appears to be a potentially useful tool in the assignment of structural class from ROA data alone. Since PVX and NMV cluster very closely together, they probably have similar coat protein folds which maybe similar to the helix bundles of the coat proteins of nearby TMV.

## 8.4 Conclusion

Principal component analysis is a powerful analytical technique that allows analysts to express their data in a way that provides a new insight into the nature of that data. Being able to look at the whole data set and compare individual members against all others simultaneously in a mathematically objective and rigorous manner is a great advantage over current visual inspection. Other techniques such as deconvolution require a level of pre-knowledge about the analytes that in many cases is simply not available. In most other cases, even a detailed description of the analytes is not enough to allow the experimental data pertaining to the analyte to be deconstructed in an objective and meaningful way.

However, due to the inherent complexity of protein ROA data, PCA cannot simply provide straightforward 'answers' to the difficult questions that some protein ROA spectra pose. The exact relationship between protein structural elements (in all their variety) and ROA spectral band patterns still needs to be explored by careful visual inspection. The power of PCA places it in a unique position to capitalize upon the wealth of experimental data being yielded by protein ROA spectroscopy. As has been demonstrated successfully using the raw PCA treatment of ROA data, PCA is able to find agreement with the major conclusions yielded by visual inspection pertaining to determination of structural class. Initial studies using trained PCA hint at the possibility of being able to concentrate on individual characteristics of sets of proteins.

# Conclusions and Further Work

The parallel development of instrumentation, data acquisition and experimental techniques has turned ROA spectroscopy into a powerful probe of biomolecular structure. The work presented in this thesis represents part of the largest survey of protein structure by ROA to date, and as such provides an opportunity to gain new insight into a wide range of issues relating to protein structure. The 39 protein ROA spectra presented here were selected from a set of over 50, all obtained on the same instrument and under largely the same experimental conditions from one protein to the next. The routine collection of data is an essential goal if ROA is to be demonstrated to be of more general use outside the research laboratory. This has largely been achieved and it will not be long before the first commercial ROA instruments become available.

The biggest problem faced by anyone attempting to carry out a survey of protein structure by ROA spectroscopy is that of understanding the data once it has been reliably collected. To this end, the wealth of experimental data collected in the past has proved invaluable in the assignment of ROA bands to protein structural properties. The fact that band assignments are continually under review is an epiphenomenon of the constant stream of new data being added to the data set. However, although more spectra should mean better understanding, the relationship between protein structure and ROA spectral band patterns is not as straightforward as one might hope. Indeed, the amount of apparently contradictory details and the lack of certainty regarding some important observed features means that the need to explore new methods of data analysis has never been greater. To this end, the data analysis part of this work has largely been successful, but remains at an early stage.

Simply put, the ideal goal of the analysis of a protein ROA spectrum is, as a minimum, a reliable estimate of the types of structure that exist within the protein. Secondly, a quantitative estimate of the relative proportions of secondary structure elements and assignment of fold type is a realistic proposition. In order to achieve this, the current methods of spectral analysis must be modified in order to take into consideration a number of critical factors that influence protein structure; for example

amino acid composition (the primary structure of the protein) and data from X-ray crystallography and NMR studies.

So far, the first attempts at quantitative analysis of protein ROA spectra involving the deconvolution of spectra have proved unsuccessful. However, the current development of principal component analysis has proved able to take into consideration a number of elements within the data that would otherwise be inaccessible. By training the PCA algorithm by the methods described in the last chapter, it is possible to extract common features within a data set, regardless of how well hidden they may be by other features that simultaneously appear.

By describing a set of data that includes percentages of secondary structure content, loop and turn content, PPII helix content etc. as well as an accurate assessment of fold type, hydration effects etc., the analysis of the ROA spectrum of protein may eventually result in a highly accurate picture of the protein from which it was obtained.

In the post-genomic era, there is an enormous demand for reliable and rapid protein structural determination techniques. The continuous development of ROA spectroscopic instrumentation and experimental technique suggests that ROA is capable of high throughput protein structure determination. The combination of this uniquely incisive technique and the rapid and objective analysis of the resulting data by principal component analysis promises to be of great value in the future.

# Appendices

# Appendix One

# Index of Protein ROA Spectra by Class Type

# Index of Protein ROA Spectra by Name

# Appendix Two

# Experimental Conditions Reference Table

| | Protein | Conc. mg/mL | Power W | Time hrs | Page |
|---|---|---|---|---|---|
| 1 | Aldolase | 100 | 1.40 | 21.052 | 79 |
| 2 | Amylase, alpha | 50 | 1.40 | 21.466 | 81 |
| 3 | Antifreeze protein III | 50 | 1.40 | 22.939 | 102 |
| 4 | Avidin | 67 | 1.40 | 24.085 | 54 |
| 5 | Bowman-Birk inhibitor | 100 | 1.20 | 24.000 | 96 |
| 6 | Casein, alpha-s1 | 50 | 1.40 | 6.013 | 117 |
| 7 | Casein, beta | 50 | 1.40 | 19.460 | 110 |
| 8 | Casein, kappa | 50 | 1.40 | 19.336 | 110 |
| 9 | Chymotrypsin, alpha | 100 | 1.40 | 9.359 | 63 |
| 10 | Concanavalin A | 67 | 1.15 | 43.505 | 52 |
| 11 | Creatine kinase | 76 | 1.40 | 32.359 | 47 |
| 12 | Immunoglobulin, human | 100 | 1.40 | 12.045 | 57 |
| 13 | Insulin | 100 | 1.40 | 12.262 | 45 |
| 14 | Invertase | 100 | 1.40 | 19.098 | 94 |
| 15 | Lactalbumin, alpha | 100 | 1.40 | 13.130 | 42 |
| 16 | Lactoglobulin, beta | 100 | 1.40 | 20.059 | 70 |
| 17 | Lysozyme, equine | 70 | 1.40 | 10.000 | 42 |
| 18 | Lysozyme, hen | 100 | 1.40 | 10.025 | 41 |
| 19 | Lysozyme, human | 50 | 1.40 | 29.373 | 41 |
| 20 | Metallothionein | 67 | 1.40 | 24.879 | 99 |
| 21 | Orosomucoid, bovine | 100 | 1.40 | 23.121 | 101 |
| 22 | Ovalbumin, hen | 100 | 1.40 | 13.279 | 85 |
| 23 | Ovomucoid, hen | 100 | 1.40 | 12.535 | 87 |
| 24 | Ovomucoid, turkey | 100 | 1.40 | 13.884 | 87 |
| 25 | Phosvitin | 100 | 1.25 | 19.263 | 116 |
| 26 | Ribonuclease A | 100 | 1.40 | 12.192 | 89 |
| 27 | Ribonuclease B | 70 | 1.40 | 12.379 | 89 |
| 28 | Serum albumin, human | 100 | 1.40 | 13.253 | 37 |
| 29 | Serum albumin, bovine | 100 | 1.40 | N/A | 37 |
| 30 | Serum amyloid P component | 56 | 1.40 | 25.886 | 59 |
| 31 | Subtilisin Carlsberg | 100 | 1.40 | 17.855 | 83 |
| 32 | Synuclein, alpha | 79 | 1.40 | 53.100 | 112 |
| 33 | Synuclein, beta | 100 | 1.40 | N/A | 112 |
| 34 | Synuclein, gamma | 100 | 1.40 | N/A | 112 |
| 35 | Tau46 wild-type | 30 | 1.40 | 21.198 | 114 |
| 36 | Tau46 P301L mutant | 30 | 1.40 | N/A | 114 |
| 37 | Trypsin | 100 | 1.40 | 19.978 | 65 |
| 38 | Trypsinogen | 100 | 1.40 | 18.976 | 67 |
| 39 | Ubiquitin | 100 | 1.25 | 21.348 | 72 |

# Appendix Three

## Amino Acid Nomenclature Table

| Name | 3-letter code | Single letter code |
|---|---|---|
| Alanine | Ala | A |
| Arginine | Arg | R |
| Asparagine | Asn | N |
| Aspartic acid | Asp | D |
| Cysteine | Cys | C |
| Glutamine | Gln | Q |
| Glutamic acid | Glu | E |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Leucine | Leu | L |
| Lysine | Lys | K |
| Methionine | Met | M |
| Phenylalanine | Phe | F |
| Proline | Pro | P |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |
| Valine | Val | V |

# Appendix Four

# Publications Arising From Work Presented In This Thesis

I    "Solution structure of native proteins with irregular folds from Raman optical activity"; Smyth, E.; Syme, C. D.; Blanch, E. W.; Hecht, L.; Vašák, M.; Barron, L. D., *Biopolymers*, **58**, 138-151 (2001)


II   "A Raman optical activity study of rheomorphism in caseins, synucleins and tau: New insight into the structure and behaviour of natively unfolded proteins"; Syme, C. D.; Blanch, E. W.; Holt, C.; Jakes, R.; Goedert, M.; Hecht, L.; Barron, L. D., *Eur. J. Biochem.*, **269**, 148-156 (2002)


III  "Solution structures of potato virus X and narcissus mosaic virus from Raman optical activity"; Blanch, E. W.; Robinson, D. J.; Hecht, L.; Syme, C. D.; Nielsen, K.; Barron, L. D., *J. Gen. Virology*, **83**, 241-246 (2002)


IV   "Molecular structure of viruses from Raman optical activity"; Blanch, E. W.; Hecht, L.; Syme, C. D.; Nielsen, K.; Barron, L. D., *Virology*, submitted (2002)

Edward Smyth[1]
Christopher D. Syme[1]
Ewan W. Blanch[1]
Lutz Hecht[1]
Milan Vašák[2]
Laurence D. Barron[1]
[1] Chemistry Department,
University of Glasgow,
Glasgow G12 8QQ, UK

[2] Biochemisches Institut der
Universität Zürich,
Winterthurerstrasse 190,
CH-8057 Zürich, Switzerland

# Solution Structure of Native Proteins with Irregular Folds from Raman Optical Activity

**Abstract:** Raman optical activity (ROA) spectra have been measured for the proteins hen phosvitin, yeast invertase, bovine α-casein, soybean Bowman–Birk protease inhibitor, and rabbit $Cd_7$-metallothionein, all of which have irregular folds in the native state. The results show that ROA is able to distinguish between two types of disorder. Specifically, invertase, α-casein, the Bowman–Birk inhibitor, and metallothionein appear to possess a "static" type of disorder similar to that in disordered states of poly(L-lysine) and poly(L-glutamic acid); whereas phosvitin appears to possess a more "dynamic" type of disorder similar to that in reduced (unfolded) lysozyme and ribonuclease A and also in molten globule protein states. In the delimiting cases, static disorder corresponds to that found in loops and turns within native proteins with well-defined tertiary folds that contain sequences of residues with fixed but nonrepetitive φ,ψ angles; and dynamic disorder corresponds to that envisaged for the model random coil in which there is a distribution of Ramachandran φ,ψ angles for each amino acid residue, giving rise to an ensemble of interconverting conformers. In both cases there is a propensity for the φ,ψ angles to correspond to the α, β and poly(L-proline) II (PPII) regions of the Ramachandran surface, as in native proteins with well-defined tertiary folds. Our results suggest that, with the exception of invertase and metallothionein, an important conformational element present in the polypeptide and protein states supporting the static type of disorder is that of the PPII helix. Long sequences of relatively unconstrained PPII helix, as in α-casein, may impart a plastic (rheomorphic) character to the structure. © 2001 John Wiley & Sons, Inc. Biopoly 58: 138–151, 2001

**Keywords:** Raman optical activity; vibrational optical activity; disordered polypeptides; disordered proteins; protein loop structure; poly(L-proline) II helix

*138*

## INTRODUCTION

Proteins with irregular structures are usually encountered in the context of the non-native states important in protein folding, stability, and function. The term "non-native" or "denatured" state embraces a plethora of structures ranging from the ideal extended random coil at one extreme[1] to collapsed molten globules with residual secondary structure and partial tertiary fold at the other.[2] However, proteins with irregular structures are also known to exist under physiological conditions and to have a biological function.[3,4] Such proteins, which can sometimes be completely unfolded, could be described as native proteins with irregular folds.

The heterogeneity of irregular protein structures, non-native or native, has made their detailed characterization difficult. A useful technique for studying such structures in aqueous solution is Raman optical activity (ROA), which measures vibrational optical activity by means of a small difference in the intensity of Raman scattering from chiral molecules in right and left circularly polarized incident laser light. ROA has recently been applied to biomolecules and provides a new perspective on their solution structure and dynamics.[5–8] ROA bears the same relation to conventional Raman spectroscopy as does uv CD to conventional uv absorption spectroscopy. ROA is able to cut through the complexity of the conventional vibrational spectra of biopolymers since the largest ROA signals are often associated with vibrational coordinates that sample the most rigid and chiral parts of the structure. In proteins these are usually within the peptide backbone and often give rise to ROA band patterns characteristic of the backbone conformation, unlike the parent Raman spectra in which many bands from the amino acid side chains often obscure the peptide backbone bands. As well as peptide backbone bands arising from secondary structure, protein ROA spectra also contain distinct bands from loops and turns, and so can provide information about the tertiary fold of the peptide backbone and its changes with pH, temperature and other variables.[9–11]

Here we report the ROA spectra of a selection of native proteins with irregular folds, together with the ROA spectra of some unordered polypeptides and reduced proteins, from which it is apparent that ROA is able to discriminate between different types of disorder. To facilitate the discussion, Figure 1 presents the MOLSCRIPT diagrams[12] of the structures of those proteins discussed below for which x-ray crystal structures are available (hen lysozyme, bovine ribonuclease A, soybean Bowman–Birk protease inhibitor, and rat metallothionein).

## EXPERIMENTAL

### Proteins, Reagents, and Sample Handling

Hen egg white lysozyme (grade I), bovine ribonuclease A, poly(L-lysine) of molecular weight ~22 kDa, poly(L-glutamic acid) of molecular weight ~83 kDa, hen egg yolk phosvitin, baker's yeast invertase (grade VII), and soybean Bowman–Birk protease inhibitor were supplied by Sigma-Aldrich Company, Ltd., and bovine milk α-casein by Fluka (Sigma-Aldrich Company, Ltd.), and used without further purification. The sample of $Cd_7$-metallothionein from rabbit liver was prepared as described previously.[13] The buffer solutions were prepared using Analar grade chemicals and distilled and deionized water. The buffers, concentrations, etc., for each sample are given in the corresponding figure captions.

Unfolded lysozyme was prepared by reducing all four disulfide bonds using the method of Creighton,[14] modified at the purification stage to dialysis six times against 3 L of 10 mM HCl (pH 2.0). Disulfide reduction is reversible so the free sulfhydryl groups are usually carboxymethylated to prevent reoxidation. However, we found that the resulting carboxymethylated protein was insufficiently soluble for ROA measurements, so we left the sulfhydryl groups unblocked and maintained a low pH to prevent reoxidation as was done in an NMR study of denatured lysozyme.[15] The solution was lyophylized and finally prepared for ROA measurements by dissolving in citrate buffer at pH 2.0. A similar procedure was followed to prepare reduced ribonuclease A except that citrate buffer at pH 2.6 was used because the sample tends to form a gel at lower pH. The number of free sulfhydryl groups was checked for both proteins using the Ellman assay,[14] and found to be between seven and eight per molecule before and after ROA data acquisition.

The polypeptide and protein solutions were made up at concentrations ~50–100 mg/mL in small glass sample tubes, mixed with a little charcoal to remove traces of fluorescing impurities, and centrifuged. The solutions were subsequently filtered through 0.22 μm Millipore filters directly into rectangular quartz microfluorescence cells that were again centrifuged gently prior to mounting in the ROA instrument. Residual visible fluorescence from traces of impurities, which can give large backgrounds in Raman spectra, was allowed to "burn down" by leaving the sample to equilibrate in the laser beam for several hours before acquiring ROA data. A new procedure adopted for these samples, which we have not used previously, was to add a small amount of potassium iodide (~8 mM) to each sample since this quenches fluorescence and so helps reduce the background. The Cd(II) ions in the metallothionein sample are kinetically labile with respect to exchange, which may lead to dimerization due to the formation of S—S bonds.[16] We therefore checked the solution immediately after the ROA spectral acquisition using dynamic light scattering and found it contained 72% monomer and 28% dimer. All the ROA spectra were acquired at room temperature (~20°C)

(a) hen lysozyme

(b) bovine ribonuclease A

(c) soybean Bowman-Birk inhibitor

(d) rat metallothionein
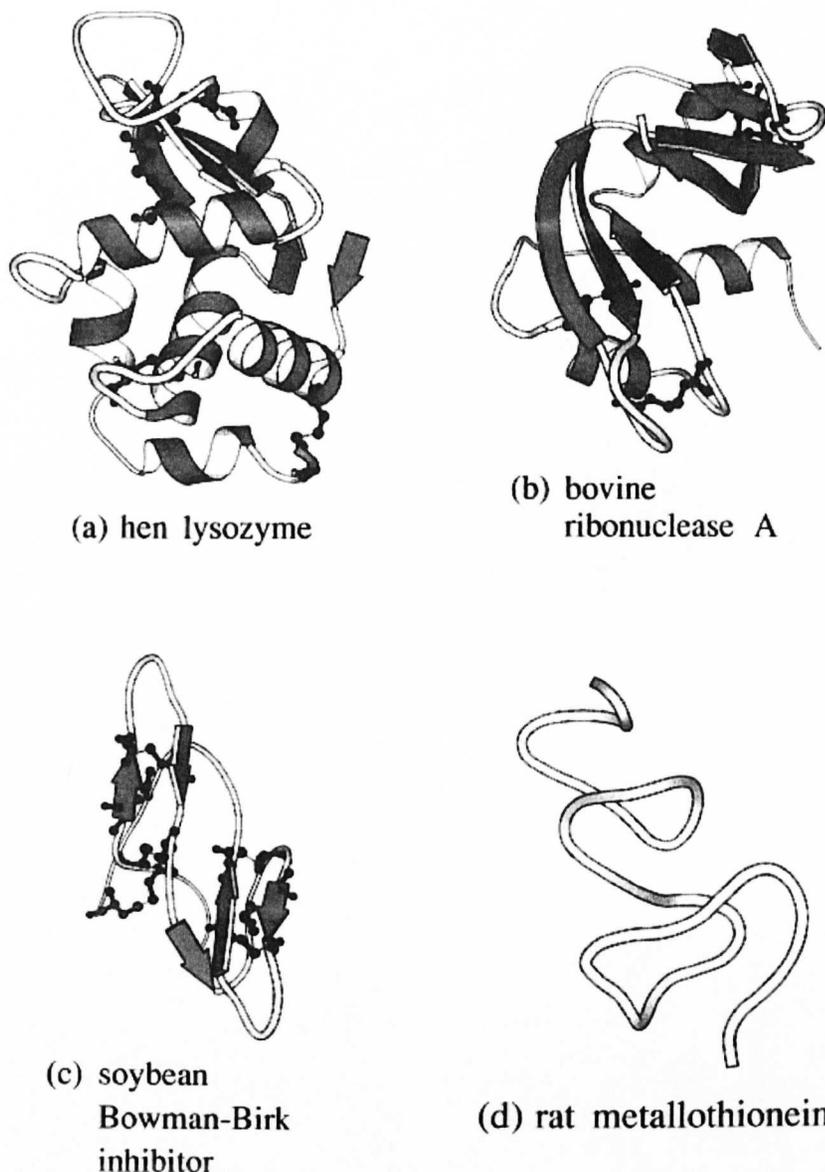
**FIGURE 1** MOLSCRIPT diagrams of the x-ray crystal structures of (a) hen lysozyme (PDB code 1lse), (b) bovine ribonuclease A (PDB code 1rbx), (c) soybean Bowman–Birk protease inhibitor (PDB code 1pi2), and (d) rat metallothionein (PDB code 4mt2).

except for that of $\alpha$-helical poly(L-lysine), which must be prepared and kept at ~0 °C to avoid gel formation.[17]

## ROA Spectroscopy

The instrument used for the Raman and ROA measurements has a backscattering configuration, which is essential for aqueous solutions of biopolymers, and employs a single-grating spectrograph fitted with a backthinned CCD camera as detector and a holographic notch filter to block the

Rayleigh line.[8] ROA is measured by synchronizing the Raman spectral acquisition with an electrooptic modulator that switches the polarization of the incident argon-ion laser beam between right- and left-circular at a suitable rate. The spectra are displayed in analog-to-digital counter units as a function of the Stokes wavenumber shift with respect to the exciting laser wavenumber. The ROA spectra are presented as circular intensity differences $I^R - I^L$ and the parent Raman spectra as circular intensity sums $I^R + I^L$, where $I^R$ and $I^L$ are the Raman-scattered intensities in right- and
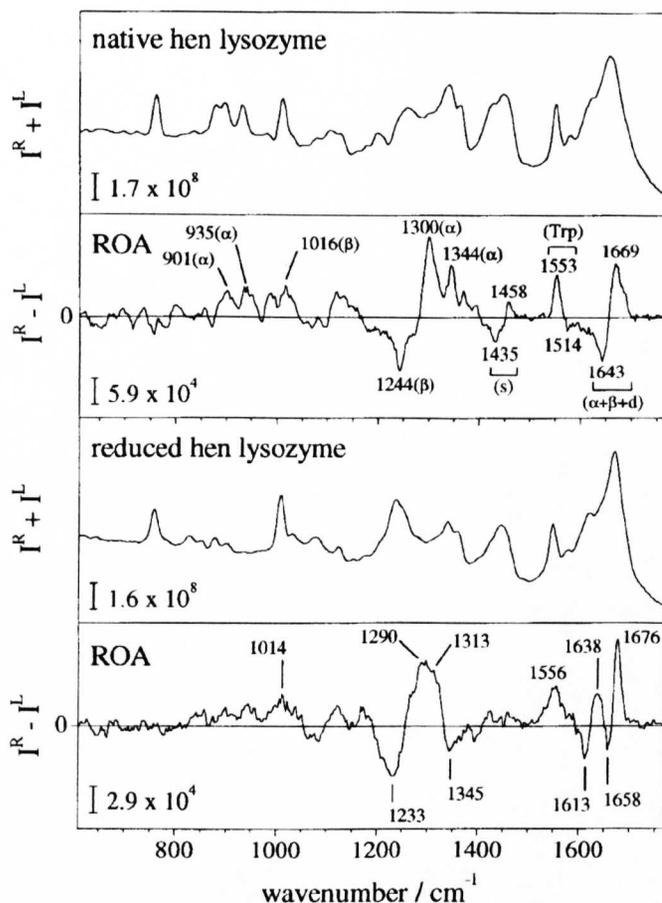
**FIGURE 2**   The backscattered Raman ($I^R + I^L$) and ROA ($I^R - I^L$) spectra of native hen lysozyme in 100 m$M$ acetate buffer at pH 5.4 (top pair) and of reduced hen lysozyme in dilute HCl at pH 2.3 (bottom pair). The structural assignments given in brackets are as follows: $\alpha$, $\alpha$ structure; $\beta$, $\beta$ structure; p, PPII structure; h, hairpin bend; s, side chain; Trp, tryptophan; d, disordered.

left-circularly polarized incident light, respectively. The experimental conditions for each sample were as follows: laser wavelength 514.5 nm; laser power at the sample ~700 mW; spectral resolution ~10 cm$^{-1}$; acquisition time ~10–20 h. For the ROA measurements at ~0°C on $\alpha$-helical poly(L-lysine), dry temperature-controlled air was blown over the sample cell from an FTS Systems model TC-84 Airjet Crystal Cooler.

## RESULTS AND DISCUSSION

### Unfolded Proteins and Disordered Polypeptides

Figures 2 and 3 show the backscattered Raman and ROA spectra of two relatively disordered non-native proteins, unfolded hen lysozyme and unfolded bovine ribonuclease A, respectively, prepared by reducing all

the disulfide bonds and keeping the sample at low pH, together with the ROA spectra of the corresponding native proteins, the structures of which are shown in Figures 1a and 1b. Due to the weaker ROA intensities and large fluctuations in the bulk sample, these reduced proteins are difficult samples for ROA measurements: the previously published versions,[18,19] although generally similar, are not as reliable in some details as the newly acquired spectra shown here. An account of the assignments of polypeptide and protein ROA bands (which are being continually refined as more ROA data accumulate) can be found in Ref. 7. A brief summary is given here for convenience.

Vibrations of the peptide backbone are usually associated with three main regions of the Raman spectrum: the backbone C$_\alpha$—C and C$_\alpha$—N skeletal stretch region ~870–1150 cm$^{-1}$; the extended amide
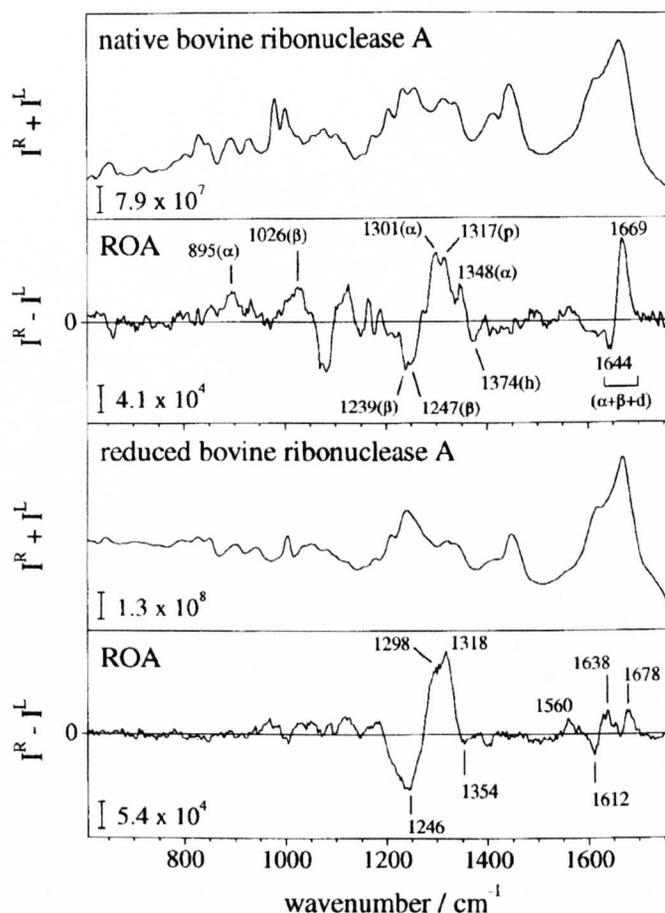
**FIGURE 3**   The backscattered Raman and ROA spectra of native bovine ribonuclease A in 100 m*M* acetate buffer at pH 5.4 (top pair) and of reduced bovine ribonuclease A in 10 m*M* citrate buffer at pH 2.4 (bottom pair).

III region ~1230–1350 cm$^{-1}$ that arises from normal modes containing various combinations of the C$_\alpha$—H and N—H deformations together with the C$_\alpha$—C and C$_\alpha$—N stretches; and the amide I region ~1630–1700 cm$^{-1}$ that arises predominantly from the C=O stretch. ROA signals in the backbone skeletal stretch region appear to arise mainly from secondary structure, those in the extended amide III region from loops and turns as well as secondary structure, and those in the amide I region mainly from secondary structure with some contributions from loops and turns. There are also a few useful ROA bands from side chains, mainly in the range ~1400–1580 cm$^{-1}$. Some ROA band assignments are indicated in the native hen lysozyme and bovine ribonuclease A spectra in Figures 2 and 3.

Since the time scale of the Raman scattering event (~3.3 × 10$^{-14}$ s for a vibration with wavenumber shift 1000 cm$^{-1}$ excited in the visible) is much shorter

than that of the fastest conformational fluctuations in biomolecules, an observed ROA spectrum is a superposition of "snapshot" spectra from all the distinct chiral conformers present in the sample at equilibrium. There are a number of significant differences in the ROA spectra of the two reduced proteins compared with those of the native states. ROA in the backbone skeletal stretch region is generally less intense in the reduced proteins than in the native states, more so in ribonuclease A than in hen lysozyme, suggesting that some secondary structure has been lost. Much of the ROA band structure in the extended amide III region has disappeared, being replaced by a large broad couplet, negative at low wavenumber and positive at high. Hints of structure are present throughout these broad couplets, suggesting that they are generated by a large number of conformations with a range of local residue Ramachandran $\phi,\psi$ angles. The envelopes of these ROA couplets are similar
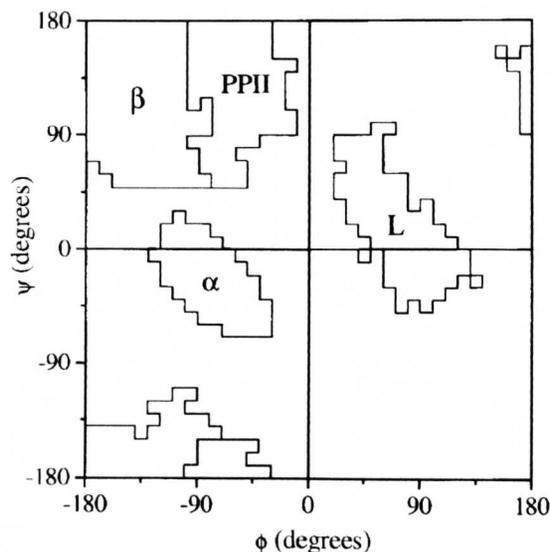
**FIGURE 4** Schematic diagram of a Ramachandran potential energy surface showing the most favored regions for residues in native proteins, namely $\alpha$, $\beta$, PPII, and L, the last standing for left-handed helix.

to those of the distinct band structures seen in the native states in the extended amide III region, which suggests that the individual residues are clustering in the same general regions of the Ramachandran surface as in the native proteins, namely those corresponding to $\alpha$, $\beta$ and poly(L-proline) II (PPII) structure illustrated in Figure 4 (adapted from Ref. 20). This is expected from other studies[1] including x-ray crystallography[20,21] and NMR.[22] Since $\beta$ structure in native proteins generates negative ROA bands at $\sim$1220 and 1245 cm$^{-1}$,[7] the broad negative components of the couplets may originate in residues clustering in the $\beta$ region of the Ramachandran surface. The broad positive components may arise from residues clustering in the $\alpha$ and PPII regions of the surface.[7] The negative ROA bands at $\sim$1350 cm$^{-1}$ may indicate the presence of local residue conformations like those in the tight turns in $\beta$ hairpins.[7] ROA structure in the range $\sim$1550–1640 cm$^{-1}$ originates in aromatic side-chain vibrations.[7] Compared with the native states the amide I couplets, negative at low wavenumber and positive at high, have shifted by $\sim$10 cm$^{-1}$ to higher wavenumber, and in the case of reduced lysozyme, this couplet has become quite sharp. In fact, the positive peaks of these amide I couplets occur at a similar wavenumber ($\sim$1676 cm$^{-1}$) to those of the positive components of the amide I couplets shown by $\beta$-sheet proteins,[7] which may reflect the fact that open turn conformations are similar to $\beta$-strand.[23] However, they are also at a

similar wavenumber to those of the positive amide I ROA bands in disordered poly(L-lysine) and poly(L-glutamic acid), and so may also be characteristic of PPII structure (vide infra). Although of similar appearance overall, the ROA spectra of reduced hen lysozyme and reduced bovine ribonuclease A displayed in Figures 2 and 3 are clearly different in detail, which may reflect the different residue compositions and their different $\phi,\psi$ propensities. In conventional Raman spectroscopy, bands in the range $\sim$1245–1270 cm$^{-1}$ in the extended amide III region and $\sim$1655–1665 cm$^{-1}$ in the amide I region are assigned to irregular structure,[24] which accords with at least some of the corresponding ROA bands.

These reduced proteins are not expected to be fully disordered (random coils) since chemical or thermal denaturation is required to induce complete unfolding. Unfortunately, it has not been possible so far to obtain ROA data on fully unfolded proteins: the use of chemical denaturants such as urea or guanadinium hydrochloride at high concentrations precludes ROA measurements due to intense Raman bands from the denaturants, and thermally unfolded proteins often show considerable light scattering due to the formation of aggregates.

Poly(L-lysine) at acidic and neutral pH and poly(L-glutamic acid) at neutral and alkaline pH have charged side chains that repel each other, thereby encouraging a disordered structure; whereas poly(L-lysine) at alkaline pH and poly(L-glutamic acid) at acidic pH have neutral side chains, and so are able to support $\alpha$-helical conformations stabilized both by internal hydrogen bonds and by hydrogen bonds to the solvent.[25] The backscattered Raman and ROA spectra of these samples in both disordered and $\alpha$-helical states are shown in Figure 5. The newly acquired ROA spectra of the two poly(L-lysine) samples shown here are similar to ones published previously.[17] There are sufficiently large differences between the ROA spectra of the corresponding disordered and $\alpha$-helical conformations to enable ROA to distinguish between these two states. The ROA spectra of the disordered polypeptides show more structure than those of the reduced proteins (Figure 2), especially in the extended amide III region, which suggests the presence of a repertoire of well-defined conformations, rather than a more dynamic structure as in the reduced proteins where there will be a distribution of $\phi,\psi$ angles for each residue, giving rise to an ensemble of conformers. Since these disordered polypeptides may contain significant amounts of extended left-handed PPII-helical structures,[26–29] it is possible that the strong positive ROA band shown by both samples at $\sim$1320 cm$^{-1}$ is one signature of PPII structure. Proteins con-
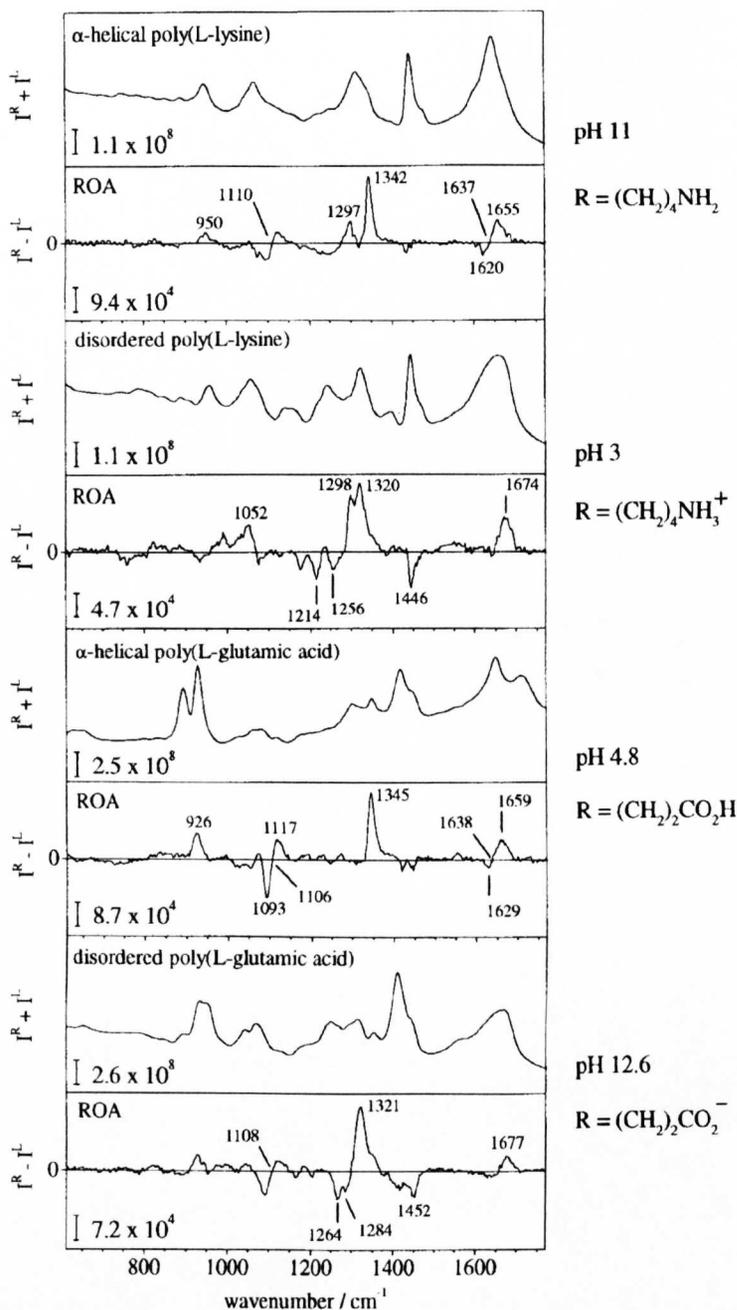
**FIGURE 5** The backscattered Raman and ROA spectra of poly(L-lysine) in α-helical (top pair) and disordered (second pair) conformations, and of poly(L-glutamic acid) in α-helical (third pair) and disordered (bottom pair) conformations in aqueous solution.

taining a significant amount of β-sheet often show a sharp positive ROA band in this region, which may originate in PPII residues within some of the longer loops connecting the β-strands or other elements of secondary structure.[7] However, there appears to be no β structure itself in these disordered polypeptides since the characteristic negative ROA bands at ~1220 and 1245 cm⁻¹ are absent; the negative ROA bands in the range ~1255–1285 cm⁻¹ may be characteristic of PPII or some related turn structure. Disordered poly(L-lysine), but not poly(L-glutamic acid), shows a positive ROA band at ~1298 cm⁻¹, suggesting the
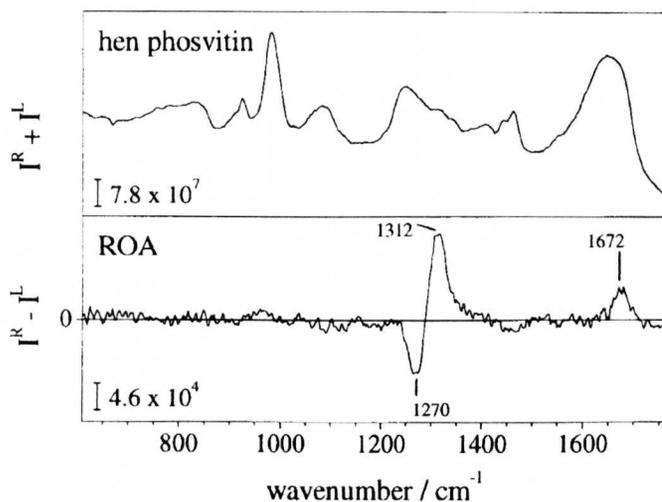
**FIGURE 6** The backscattered Raman and ROA spectra of hen phosvitin in 100 m$M$ acetate buffer at pH 5.4.

presence of some $\alpha$ structure. Both disordered polypeptides show weak broad positive ROA bands in the amide I region at ~1675 cm$^{-1}$. The general appearance of the ROA spectra of disordered polypeptides therefore suggests that they have structures akin to those of some of the longer loops in native folded proteins: these contain sequences of residues with fixed but nonrepetitive $\phi,\psi$ angles, most of which are located in the same regions of the Ramachandran surface as are most heavily populated in native proteins.[30,31]

## Phosvitin

Phosvitin is a highly hydrophilic phosphoglycoprotein that constitutes the major protein in egg yolk and represents 1% of the mass.[32] The variant found in hen egg yolk has a molecular weight of ~34 kDa with carbohydrate constituting ~6.5% of the molecular mass. The sequence of hen phosvitin contains 216 residues and is unusual in that 123 of these are serines, most of which are phosphorylated. The three-dimensional structure of phosvitin has not been determined, and spectroscopic studies have suggested several different possibilities for the solution conformation. At neutral pH, NMR,[33] conventional Raman,[34] and vibrational CD (VCD)[35] indicate an irregular structure, whereas uv CD and Fourier transform infrared[36] indicate the presence of some $\beta$-sheet in addition to irregular structure. At low pH, $\beta$-sheet appears to predominate.[34,35]

The backscattered Raman and ROA spectra of hen egg phosvitin in aqueous solution are shown in Figure

6. It is immediately apparent from several independent regions of the ROA spectrum that virtually no secondary structure is present. In particular, no significant ROA signals appear in the backbone skeletal stretch region, and there is no sign of a couplet, negative at low wavenumber and positive at high, in the amide I region characteristic of $\alpha$-helix or $\beta$-sheet.[7] However, the small positive ROA band at ~1675 cm$^{-1}$ is similar to those seen in the two reduced proteins and the two disordered polypeptides above, and so may also arise from open turn conformations similar to $\beta$-strand or PPII structure. The couplet, with a negative peak (possibly structured) at ~1270 cm$^{-1}$ and a positive peak (again possibly structured) at ~1312 cm$^{-1}$, in the extended amide III region is unusually narrow, which suggests a smaller range of individual residue $\phi,\psi$ than those in the reduced proteins and may reflect the predominance of phosphorylated serines. It has been suggested previously from VCD data that phosvitin adopts similar conformations to those in disordered poly(L-lysine) and poly(L-glutamic acid), especially PPII structure.[35] However, at ~1312 cm$^{-1}$ the positive ROA peak seems a little low to be assigned to PPII structure. Unfortunately, we were unable to measure the ROA spectrum of phosvitin at low pH, under which conditions it is thought to take up a $\beta$-sheet conformation,[34,35] due to gel formation.

## Invertase

Yeast external invertase is a biocatalyst used in the sugar industry. It is a globular protein of molecular
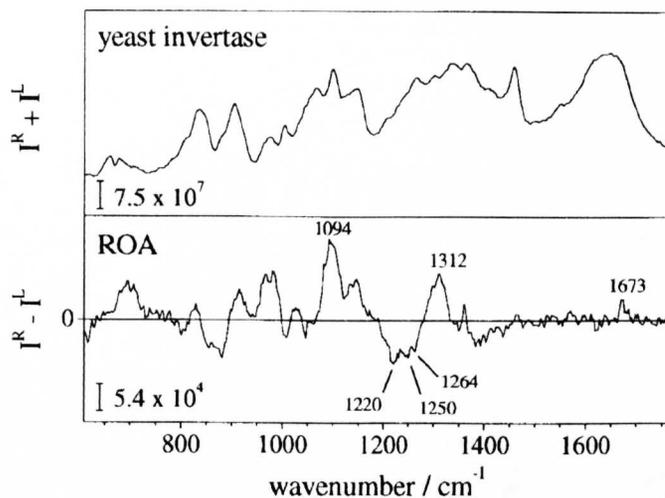
**FIGURE 7**  The backscattered Raman and ROA spectra of yeast invertase in 100 m*M* acetate buffer at pH 5.4.

weight ~270 kDa encompassing two subunits with up to 50% of its mass as carbohydrate in the form of nine high-mannose oligosaccharide chains, each linked to an asparagine residue on the polypeptide backbone. No x-ray crystal structure or NMR solution structure is available. Spectroscopic evidence is contradictory. The uv CD spectrum of an aqueous solution indicates little if any $\alpha$-helix,[37] but could be interpreted as indicating a significant amount of $\beta$-sheet, which structure prediction methods suggest is a $\beta$-propeller fold[38]; however, a conventional Raman study suggests a predominance of $\alpha$-helix in aqueous solution and $\beta$-sheet in the solid state.[39]

The backscattered Raman and ROA spectra of yeast external invertase in aqueous solution are shown in Figure 7. The absence of an ROA couplet in the amide I region would seem to provide strong evidence that this protein contains virtually no $\alpha$-helix or $\beta$-sheet; only a very weak positive ROA band is present in this region. It is therefore perplexing that the presence of some $\beta$-sheet might be inferred from the negative ROA bands at ~1220 and 1250 cm$^{-1}$. One possible explanation is that these two $\beta$-sheet ROA signatures reflect the local residue $\phi,\psi$ angles within $\beta$-strands irrespective of whether or not the strands are associated via hydrogen bonds within $\beta$ sheet, whereas the characteristic amide I ROA couplet (which is expected to be sensitive to interstrand hydrogen bonding since it is largely due to C=O stretch coordinates) is only generated if the interstrand hydrogen bonds are present. In which case yeast external invertase may contain unassociated $\beta$-strands, perhaps stabilized by the large amount of carbohydrate. The positive ROA band at ~1312 cm$^{-1}$ with a broad

shoulder on the low-wavenumber side may originate in turn structure, as in metallothionein (vide infra).

Most of the strong ROA bands below ~1150 cm$^{-1}$ probably arise mainly from the mannose-based oligosaccharide chains, but we have not assigned them since mannose and its associated oligosaccharides have yet to be characterized using ROA. A conventional Raman band at 1081 cm$^{-1}$ has been assigned to peripheral phosphates from oligosidic chains,[39] so the possibility that the strong positive ROA band at ~1094 cm$^{-1}$ may originate in vibrations of the phosphate groups should also be kept in mind.

## $\alpha$-Casein

The casein proteins constitute nearly 80% of bovine milk. The major components, $\alpha$- and $\beta$-casein, have molecular weights ~24,000 but form aggregates with molecular weights up to ~500,000.[40,41] The monomers contain only one disulfide link[42] and so are relatively unconstrained structures. Early spectroscopic work suggested that caseins are largely "structureless" with little extended secondary structure, but later uv CD studies suggested that, although largely "random coil," $\alpha$- and $\beta$-casein may contain ~20% $\alpha$-helix and possibly a small amount of $\beta$ sheet.[40] A conventional Raman study indicated ~10% $\alpha$-helical structure and 20% $\beta$ structure in both $\alpha$- and $\beta$-casein, but the different fine structure in the two Raman spectra suggested that their conformations are not identical.[41] Sequence-based structure prediction methods suggest that the caseins are of the all $\beta$-strand type, but that condensation into $\beta$-sheets is
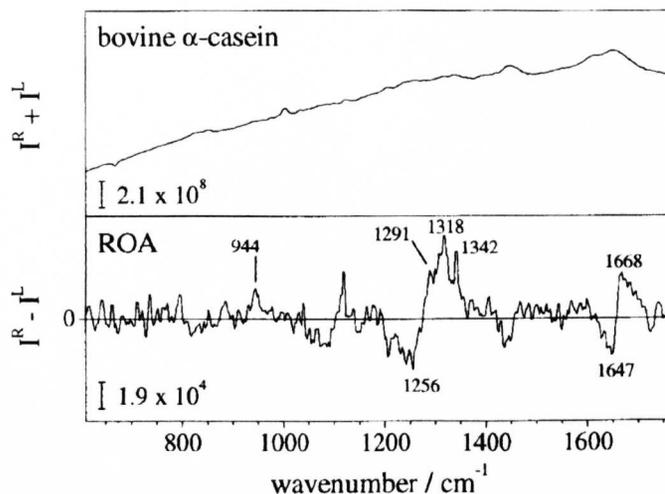
**FIGURE 8**  The backscattered Raman and ROA spectra of bovine α-casein in 100 m$M$ acetate buffer at pH 5.4.

inhibited allowing the proteins to maintain an open and mobile "rheomorphic" (flowing shape) conformation.[42]

The backscattered Raman and ROA spectra of bovine α-casein are shown in Figure 8. Although, due to the high fluorescence background of our poor quality commercial sample, the signal-to-noise ratio of the ROA spectrum is not as good as the others shown in this paper, some useful information can nonetheless be deduced. It is clear that this irregular native protein contains more well-defined structure than the previous two. The positive ROA band at ~944 cm$^{-1}$ together with the amide I couplet, negative at ~1647 and positive at ~1668 cm$^{-1}$, are characteristic of the α-helix, which appears to be hydrated on account of the sharp positive ROA band at ~1342 cm$^{-1}$.[7] Furthermore, the strong positive ROA band at ~1318 cm$^{-1}$ suggests the presence of much PPII structure. Judging by the negative ROA intensity in the range ~1240–1250 cm$^{-1}$, there may also be some β structure present. A more definitive ROA characterization of the caseins must await the availability of higher purity samples.

## Bowman–Birk Protease Inhibitor

Bowman–Birk protease inhibitors are small proteins of molecular weight ~7–9 kDa, present in the seeds of many leguminous plants. A large number have been described and sequenced.[43,44] They all have the unusually large number of seven disulfide links, which hold this small single chain protein in its well-defined native fold comprising two tandem homologous domains, each containing one protease binding site. They are of considerable current interest for medical and agricultural applications[44] on account of, in addition to antiprotease activity, their anticarcinogenic activity, immune-stimulating properties, and the ability to defend plants from insect attack. They also have high resistance to acids and heat treatments. The x-ray crystal structure of isoform PI-II of the soybean variant has been reported (PDB code 1pi2),[45] and according to the CATH system,[46] is classed as having few secondary structures with an architecture that is irregular. According to 1pi2, there are 61 residues in the chain, which comprises five short stretches of β-strand contained in two small sections of antiparallel sheet, as illustrated in the MOLSCRIPT diagram in Figure 1c. A solution NMR structure of isoform BBI-I of the soybean variant, which contains 71 residues, is also available and is generally similar to the x-ray crystal structure of isoform PI-II.[47]

The backscattered Raman and ROA spectra of the soybean Bowman–Birk inhibitor are shown in Figure 9. Despite its irregular x-ray crystal structure, the ROA spectrum is qualitatively quite different from those of phosvitin and invertase in that it contains sharp detailed structure throughout, as found in native proteins with more regular folds. In fact, a solution uv CD investigation[44] suggested that a significant amount of secondary structure is present (the best fit of structural parameters to the experimental uv CD spectrum was 9% helical, 41% β-sheet, 22% β-turn, and 28% random coil). The strong sharp negative extended amide III ROA band at ~1239 cm$^{-1}$ is consistent with a significant content of well-defined
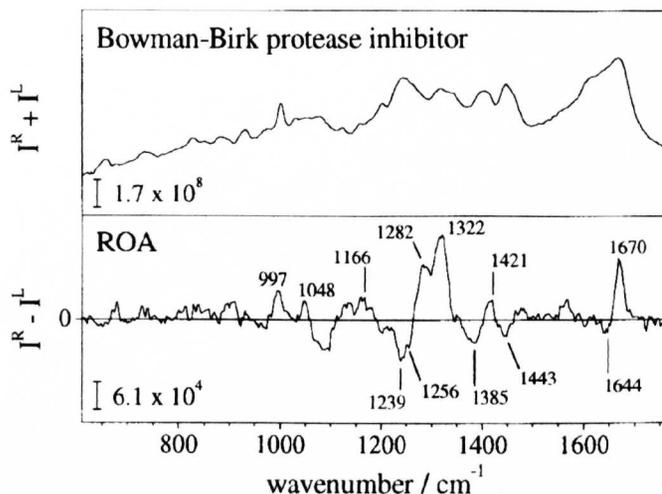
**FIGURE 9**  The backscattered Raman and ROA spectra of soybean Bowman–Birk protease inhibitor in 50 m$M$ acetate buffer at pH 5.4.

$\beta$-strand, and the very strong sharp positive band at ~1322 cm$^{-1}$ with a large amount of residues with PPII structure within the long loops. According to 1pi2, and also the solution NMR structure,[47] there are indeed a large number of residues with $\phi,\psi$ angles distributed approximately uniformly over the contiguous $\beta$ and PPII regions of the Ramachandran surface. The appearance of the amide I couplet with its strong sharp positive peak at ~1670 cm$^{-1}$ is also consistent with these assignments.[7] The broad negative ROA band centered at ~1385 cm$^{-1}$ is consistent with the presence of the two $\beta$-hairpins specified in 1pi2.[7] The positive ROA peak at ~1282 cm$^{-1}$ may originate in residues in loops and turns with $\phi,\psi$ angles within the $\alpha$ region of the Ramachandran surface (vide infra).[7,11] Overall, the ROA spectrum gives the impression of a constrained structure containing residues with $\phi,\psi$ angles fixed in well-defined regions of the Ramachandran surface.

## Metallothionein

Metallothioneins are small cysteine-rich proteins, molecular weight ~6 kDa, that have the ability to accommodate heavy metal ions such as $Zn^{2+}$ and $Cd^{2+}$.[48] They are ubiquitous tissue components in higher organisms, and have been tentatively attributed both unspecific protective functions against toxic metal ions and specific roles in zinc-regulated cellular processes. The amino acid sequence of 61 or 62 residues is completely void of aromatic residues. Rather than forming disulfide bridges, all 20 cysteine residues bind seven divalent metal ions that, among

other things, serves to stabilize the native fold of this small protein.[48] Both x-ray crystal[49] and solution NMR[50] structures are available for the rat variant and are very similar.[48] Additionally, solution NMR structures are available for the rabbit[13] and human variants.[51] All reveal sequences of type I and type II turns with no extended secondary structure except for small amounts of $3_{10}$-helix, the largest being ~10% in human metallothionein.[51] A MOLSCRIPT diagram of the x-ray structure of the rat variant, PDB code 4mt2, is shown in Figure 1d. A study using conventional Raman spectroscopy reported results consistent with a predominantly turn-containing structure.[52]

The backscattered Raman and ROA spectra of rabbit Cd$_7$-metallothionein are shown in Figure 10. Like the Bowman–Birk inhibitor, the ROA spectrum contains sharp structure throughout, although there are fewer bands than in the Bowman–Birk protein, suggesting fewer distinct structural types. In particular, there are no negative ROA bands in the range ~1220–1260 cm$^{-1}$ in the extended amide III region, which is consistent with the absence of $\beta$ structure, and no couplet characteristic of $\beta$-sheet appears in the amide I region, only a broad positive band characteristic of disordered structure. The strong positive ROA band peaking at ~1312 cm$^{-1}$ with a broad shoulder extending down to ~1275 cm$^{-1}$ on the low-wavenumber side may originate in residues within the turns with $\phi,\psi$ angles residing in the general $\alpha$ region of the Ramachandran surface, which is indeed significantly populated according to the Ramachandran plot of the PDB X-ray structure 4mt2 of the rat variant (the PDB NMR structures, e.g., 1mrb for the rabbit variant
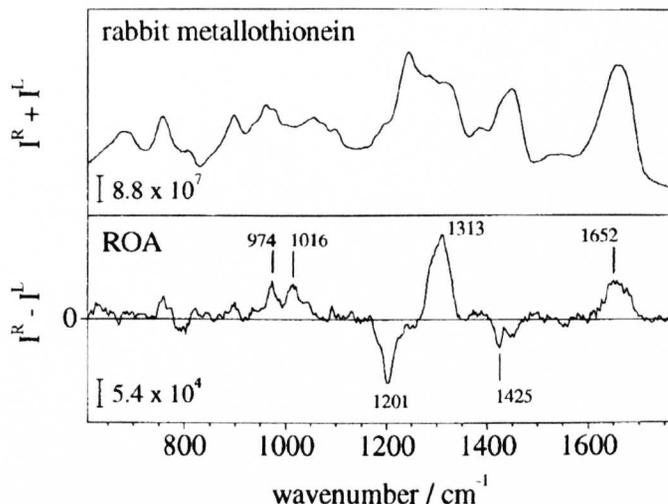
**FIGURE 10** The backscattered Raman and ROA spectra of rabbit metallothionein in 50 m$M$ Tris buffer at pH 8.0.

studied here, are of lower resolution and are less well refined and accordingly show a greater scatter of residues over their respective Ramachandran plots). Structures within this general $\alpha$ region include $3_{10}$-helix and type I $\beta$ turns. This is also the wavenumber range to which $\beta$-turns have been assigned in conventional Raman spectroscopy.[53] Analogous to the possibility of unassociated $\beta$-strands in invertase, the absence of a characteristic ROA $\alpha$-helix couplet in the amide I region for metallothionein could be due to the absence of intrachain $\alpha$-helix hydrogen bonds, with positive ROA in the $\sim$1300 cm$^{-1}$ region simply reflecting local residue $\phi,\psi$ angles within the $\alpha$ region irrespective of whether or not the residues are part of extended $\alpha$-helix.

The strong sharp negative ROA band at $\sim$1201 cm$^{-1}$ is most unusual since little ROA intensity is seen here in most proteins. This band may be associated with the large number of cysteine residues present in the protein, a suggestion supported by the presence of a large negative band at a similar wavenumber in the ROA spectrum of a concentrated solution ($\sim$100 mg/mL) of the tripeptide glutathione ($\gamma$-Glu–Cys–Gly) in H$_2$O (data not shown), and also in that of cysteine itself.[54]

Since aromatic side chains increase the population of $\beta$ space,[1] the complete absence of aromatic residues from metallothionein accords with the small number of residues in the general $\beta$ region of the Ramachandran plot, according to PDB structure 4mt2, and the absence of bands characteristic of $\beta$ structure in the ROA spectrum of Figure 10. Like that of the Bowman–Birk inhibitor, the ROA spectrum overall suggests a constrained structure.

## CONCLUSIONS

This study of a range of native proteins with irregular folds has revealed that ROA is able to sense two distinct types of disorder in aqueous solution. The delimiting cases are a "static" type of disorder in which each residue populates only a single $\phi,\psi$ conformation with small fluctuations around the mean angles; and a "dynamic" type of disorder in which there is a distribution of $\phi,\psi$ angles for each amino acid residue giving rise to an ensemble of interconverting conformers, as envisaged for the model random coil.[1] In both cases there is a propensity for the $\phi,\psi$ angles to correspond to the $\alpha$, $\beta$, and PPII regions of the Ramachandran surface, as in native proteins with well-defined tertiary folds.

Of the samples studied here, invertase, $\alpha$-casein, the Bowman–Birk inhibitor, and metallothionein appear to support disorder of the static type similar to that in disordered states of poly(L-lysine) and poly(L-glutamic acid), and phosvitin supports a more dynamic type of disorder akin to that in reduced lysozyme and ribonuclease A. Apart from evidence of more secondary structure, the ROA spectra of molten globule protein states, such as that of $\alpha$-lactalbumin at low pH,[7–9] often look similar to those of the two reduced proteins, which indicates that they too contain dynamically disordered sequences.

Since some of the ROA bands of native proteins assigned to loop and turn structure also appear in the ROA spectra of polypeptide and protein states supporting the static type of disorder, it appears that static disorder is akin to that found in the loops and turns within native proteins with well-defined tertiary folds

and that contain sequences of residues with fixed but nonrepetitive $\phi,\psi$ angles.[20,30,31] Our results suggest that, as monitored by a positive ROA band at ~1320 cm$^{-1}$, an important conformational element present in many polypeptide sequences supporting static disorder is that of PPII helix, observed here in disordered poly(L-lysine), poly(L-glutamic acid), $\alpha$-casein, and the Bowman–Birk inhibitor. This accords with the conclusion from VCD studies that the "random coil" state of polypeptides has locally ordered regions with a conformation similar to that of PPII helix.[27,28] The geometry of PPII helix observed in protein x-ray crystal structures, together with its special characteristics, which include lack of intrachain hydrogen bonds and considerable backbone hydration, allows the polypeptide chain to progress easily from this conformation to most other types, making it a favorable conformational element in longer loops connecting secondary structure elements.[55] Long sequences of relatively unconstrained PPII helix may therefore impart a plastic (rheomorphic) character to the overall structure of a protein, as suggested for the caseins.[42] Since the ROA spectra of invertase and metallothionein lack any positive intensity at ~1320 cm$^{-1}$, we conclude that neither contains any PPII structure. This may be due to the stabilization of other conformational elements in metallothionein by metal ion binding to the 20 cysteine residues and in invertase by the large amount of carbohydrate.

## REFERENCES

1. Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. Fold Des 1996, 1, R95–R106.
2. Dobson, C. M. Curr Biol 1994, 4, 636–640.
3. Gast, K.; Damaschun, H.; Eckert, K.; Schulze-Forster, K.; Rainer Maurer, H.; Müller-Frohne, M.; Zirwer, D.; Czarnecki, J.; Damaschun, G. Biochemistry 1995, 34, 13211–13218.
4. Penkett, C. J.; Redfield, C.; Jones, J. A.; Dodd, I.; Hubbard, J.; Smith, R. A. G.; Smith, L. J.; Dobson, C. M. Biochemistry 1998, 37, 17054–17067.
5. Barron, L. D.; Hecht, L.; Bell, A. F.; Wilson, G. Appl Spectrosc 1996, 50, 619–629.
6. Vargek, M.; Freedman, T. B.; Nafie, L. A. J Raman Spectrosc 1997, 28, 627–633.
7. Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. Prog Biophys Mol Biol 2000, 73, 1–49.
8. Hecht, L.; Barron, L. D.; Blanch, E. W.; Bell, A. F.; Day, L. A. J Raman Spectrosc 1999, 30, 815–825
9. Wilson, G.; Ford, S. J.; Cooper, A.; Hecht, L.; Wen, Z. Q.; Barron, L. D. J Mol Biol 1995, 254, 747–760.
10. Teraoka, J.; Bell, A. F.; Hecht, L.; Barron, L. D. J Raman Spectrosc 1998, 29, 67–71.
11. Blanch, E. W.; Hecht, L.; Barron, L. D. Protein Sci 1999, 8, 1362–1367.
12. Kraulis, P. J. J Appl Cryst 1991, 24, 946–950.
13. Arseniev, A.; Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. J Mol Biol 1988, 201, 637–657.
14. Creighton, T. E. In Protein Structure: a Practical Approach; Creighton, T. E., Ed.; IRL Press: Oxford, 1989; pp 155–167.
15. Evans, P. A.; Topping, K. D.; Woolfson, D. N.; Dobson, C. M. Proteins Struct Funct Genet 1991, 9, 248–266.
16. Otvos, J. D.; Engeseth, H. R.; Wehrli, S. Biochemistry 1985, 24, 6735–6740.
17. Wilson, G.; Hecht, L.; Barron, L. D. J Chem Soc Faraday Trans 1996, 92, 1503–1510.
18. Wilson, G.; Hecht, L.; Barron, L. D. Biochemistry 1996, 35, 12518–12525.
19. Ford, S. J.; Cooper, A.; Hecht, L.; Wilson, G.; Barron, L. D. J Chem Soc Faraday Trans 1995, 91, 2087–2093.
20. Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. Nature Struct Biol 1995, 2, 596–603.
21. Adzhubei, A. A.; Eisenmenger, F.; Tumanyan, V. G.; Zinke, M.; Brodzinski, S.; Esipova, N. G. J Biomol Struct Dynam 1987. 5, 689–704.
22. Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. J Phys Chem 1996, 100, 2661–2666.
23. Sundaralingam, M.; Sekharudu, Y. C. Science 1989, 244, 1333–1337.
24. Miura, T.; Thomas, G. J. In Subcellular Biochemistry, Volume 24. Proteins: Structure, Function, and Engineering; Biswas, B. B., Roy, S., Eds.; Plenum Press: New York, 1995; pp 55–59.
25. Bergethon, P. R. The Physical Basis of Biochemistry; Springer-Verlag: New York, 1998.
26. Tiffany, M. L.; Krimm, S. Biopolymers 1968, 6, 1379–1382.
27. Dukor, R. K.; Keiderling, T. A. Biopolymers 1991, 31, 1747–1761.
28. Keiderling, T. A.; Silva, R. A. G. D.; Yoder, G.; Dukor, R. K. Bioorg Med Chem 1999, 7, 133–141.
29. Woody, R. W. Adv Biophys Chem 1992, 2, 37–79.
30. Geetha, V.; Munson, P. J.; Protein Sci 1997, 6, 2538–2547.
31. Donate, L. E.; Rufino, S. D.; Canard, L. H. J.; Blundell, T. L. Protein Sci 1996, 5, 2600–2616.
32. Burley, R. W.; Vadhera, D. V. The Avian Egg: Chemistry and Biology; Wiley: New York, 1989.
33. Vogel, H. J. Biochemistry 1983, 22, 668–674.
34. Prescott, B.; Renugopalakrishnan, V.; Glimcher, M. J.; Bhushan, A.; Thomas, G. J. Biochemistry 1986, 25, 2792–2798.

35. Yasui, S. C.; Pancoska, P.; Dukor, R. K.; Keiderling, T. A.; Renugopalakrishnan, V.; Glimcher, M. J.; Clark, R. C. J Biol Chem 1990, 265, 3780–3788.

36. Losso, J. N.; Bogumil, R.; Nakai, S. Comp Biochem Physiol 1993, 106B, 919–923.

37. Schülke, N.; Schmid, F. X. J Biol Chem 1988, 263, 8832–8837.

38. Pons, T.; Olmea, O.; Chinea, G.; Beldarraín, A.; Márquez, G.; Acosta, N.; Rodríguez, L.; Valencia, A. Proteins Struct Funct Genet 1998, 33, 383–395.

39. Athès, V.; Combes, D.; Zwick, A. J Raman Spectrosc 1998, 29, 373–378.

40. Creamer, L. K.; Richardson, T.; Parry, D. A. D. Arch Biochem Biophys 1981, 211, 689–696.

41. Blyler, D. M.; Susi, H. J Indust Microbiol 1988, 3, 73–88.

42. Holt, C.; Sawyer, L. J Chem Soc Faraday Trans 1993, 89, 2683–2692.

43. Ikenaka, T.; Norioka, S. In Proteinase Inhibitors; Barrett, A. J., Salvesen, G., Eds.; Elsevier: Amsterdam, 1986; pp 361–374.

44. de la Sierra, I. L.; Quillien, L.; Flecker, P.; Gueguen, J.; Brunie, S. J Mol Biol 1999, 285, 1195–1207.

45. Chen, P.; Rose, J.; Love, R.; Wei, C. H.; Wang, B. C. J Biol Chem 1992, 267, 1990–1994.

46. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., Thornton, J. M. Structure 1997, 5, 1093–1108.

47. Werner, M. H.; Wemmer, D. E. Biochemistry 1992, 31, 999–1010.

48. Braun, W.; Vašák, M.; Robbins, A. H.; Stout, C. D.; Wagner, G.; Kägi, J. H. R.; Wüthrich, K. Proc Natl Acad Sci USA 1992, 89, 10124–10128.

49. Robbins, A. H.; McRee, D. E.; Williamson, M.; Collett, S. A.; Xuong, N. H.; Furey, W. F.; Wang, B. C.; Stout, C. D. J Mol Biol 1991, 221, 1269–1293.

50. Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. J Mol Biol 1988, 203, 251–268.

51. Messerle, B. A.; Schäffer, A.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. J Mol Biol 1990, 214, 765–779.

52. Pande, J.; Pande, C.; Gilg, D.; Vašák, M.; Callender, R.; Kägi, J. H. R. Biochemistry 1986, 25, 5526–5532.

53. Krimm, S.; Bandekar, J. Adv Protein Chem 1986, 38, 181–364.

54. Gargaro, A. R.; Barron, L. D.; Hecht, L. J Raman Spectrosc 1993, 24, 91–96.

55. Adzhubei, A. A.; Sternberg, M. J. E. J Mol Biol 1993, 229, 472–493.

# A Raman optical activity study of rheomorphism in caseins, synucleins and tau
## New insight into the structure and behaviour of natively unfolded proteins

**Christopher D. Syme[1], Ewan W. Blanch[1], Carl Holt[2], Ross Jakes[3], Michel Goedert[3], Lutz Hecht[1] and Laurence D. Barron[1]**

[1]*Department of Chemistry, University of Glasgow, UK;* [2]*Hannah Research Institute, Ayr, UK;* [3]*Medical Research Council Laboratory of Molecular Biology, Cambridge, UK*

The casein milk proteins and the brain proteins α-synuclein and tau have been described as natively unfolded with random coil structures, which, in the case of α-synuclein and tau, have a propensity to form the fibrils found in a number of neurodegenerative diseases. New insight into the structures of these proteins has been provided by a Raman optical activity study, supplemented with differential scanning calorimetry, of bovine β- and κ-casein, recombinant human α-, β- and γ-synuclein, together with the A30P and A53T mutants of α-synuclein associated with familial cases of Parkinson's disease, and recombinant human tau46 together with the tau46 P301L mutant associated with inherited frontotemporal dementia. The Raman optical activity spectra of all these proteins are very similar, being dominated by a strong positive band centred at ≈ 1318 cm$^{-1}$ that may be due to the poly(L-proline) II (PPII) helical conformation.

There are no Raman optical activity bands characteristic of extended secondary structure, although some unassociated β strand may be present. Differential scanning calorimetry revealed no thermal transitions for these proteins in the range 15–110 °C, suggesting that the structures are loose and noncooperative. As it is extended, flexible, lacks intrachain hydrogen bonds and is hydrated in aqueous solution, PPII helix may impart a rheomorphic (flowing shape) character to the structure of these proteins that could be essential for their native function but which may, in the case of α-synuclein and tau, result in a propensity for pathological fibril formation due to particular residue properties.

*Keywords*: caseins, synucleins and tau; polyproline II helix; amyloid fibrils; neurodegenerative disease; Raman optical activity.

Although nonregular protein structures are usually encountered under certain denaturing conditions, it is becoming increasingly apparent that proteins with nonregular structures also exist under physiological conditions [1]. The fact that such proteins can have important biological functions has necessitated a reassessment of the structure–function paradigm [2]. Native proteins with nonregular structures include the casein milk proteins [3], the phosphophoryns of bone and the phosvitins of egg yolk [4], Bowman–Birk protease inhibitors [5], metallothioneins [6], prothymosin α [7], a bacterial fibronectin-binding protein [8], the brain protein α-synuclein together with the related proteins β-synuclein and γ-synuclein [9–12], and the brain protein tau [13–16]. In addition to their role in normal function, nonregular protein structures in both non-native and native states are also of interest on account of their susceptibility to the type of aggregation found in many protein misfolding diseases.

The heterogeneity of nonregular protein structures, non-native or native, has made their detailed characterization difficult. As a result, all nonregular protein structures are often called random coil, implying that they behave like synthetic high polymers in dilute aqueous solution for which the random coil model was originally developed. The random coil state is envisaged as the collection of an enormous number of possible random conformations of an extremely long molecule in which chain flexibility arises from internal rotation (with some degree of hindrance) around the covalent backbone bonds [17]. However, there is a growing awareness that this extreme situation does not occur in most nonregular protein states. In order to further our understanding of the behaviour of proteins with nonregular structures, it is necessary to employ experimental techniques able to discriminate between the dynamic true random coil state and more static types of disorder.

One such technique is Raman optical activity (ROA), which measures vibrational optical activity by means of a small difference in the intensity of Raman scattering from chiral molecules in right- and left-circularly polarized incident laser light [18]. It has recently been demonstrated that ROA is able to distinguish two distinct types of disorder in nonregular protein structures in aqueous solution [19]. The delimiting cases are a dynamic disorder corresponding to that envisaged for the random coil in

which there is a distribution of Ramachandran $\phi,\psi$ angles for each residue, giving rise to an ensemble of rapidly interconverting conformers, and a static disorder corresponding to that found in loops and turns within native proteins with well-defined tertiary folds that contain sequences of residues with fixed but nonrepetitive $\phi,\psi$ angles.

A dominant conformational element present in more static types of disorder appears to be that of the left-handed poly(L-proline) II (PPII) helix [19]. Although PPII structure can be distinguished from random coil in peptides using ultraviolet circular dichroism (UVCD) [20] and vibrational circular dichroism (VCD) [21], these techniques are less sensitive than ROA for detecting PPII structure when there are a number of other conformational elements present as in proteins. As it is extended, flexible and hydrated, PPII helix imparts a plastic open character to the structure and may be implicated in the formation of regular fibrils in the amyloid diseases [22].

A distinction should be made between 'native proteins with nonregular structures' and 'natively unfolded' proteins. Both refer to proteins containing little regular secondary structure. However the latter, which are a special case of the former, are loose structures that simply become looser through a continuous transition on heating which takes them closer to the true random coil. The broader term 'native proteins with nonregular structures', on the other hand, also encompasses proteins with fixed nonregular folds stabilized by, for example, cooperative side chain interactions, multiple disulfide links or multiple metal ions. These fixed folds may often (but not always) be shown by a first-order thermal transition observed using DSC, and are sometimes accessible through X-ray crystallography.

It has already been suggested by Holt & Sawyer [3] that the open and relatively mobile conformation of the caseins, which allows rapid and extensive degradation to smaller peptides by proteolytic enzymes, is better described as rheomorphic, meaning flowing shape, than random coil. These authors also suggested that the rheomorphic conformation of the casein phosphoproteins was important in protecting the mammary gland against pathological calcification during lactation. This function depends on the ability of the protein to combine rapidly with nuclei of calcium phosphate to form stable calcium phosphate nanoclusters [23,24].

The synucleins, which are also usually described as random coil proteins [10–12], may have similar structural characteristics that could provide clues to their physiological functions, which are as yet unclear, and may also provide insight into why $\alpha$-synuclein forms the amyloid fibrils associated with Parkinson's disease and several other neurodegenerative diseases [25,26]. This is similar to the case for tau protein, which forms the filaments found in neuronal inclusions in Alzheimer's and other neurodegenerative diseases [26], except that the function of tau is known: it promotes and stabilizes the assembly of microtubules [27]. Although the caseins are not usually associated with any propensity to form amyloid fibrils, the presence of amyloid-like plaques in the proteinaceous parts of calcified stones known as corpora amylacea has recently been reported [28]. Such stones form in the mammary gland during lactation and contain a group of amyloid-staining peptides that start at position 81 of $\alpha_{S2}$-casein. In another

recent report, reduced $\kappa$-casein was observed to polymerize into long rod-like structures when heated to 37 °C [29].

In this paper, the theme of PPII structure and rheomorphism is explored by a comparative ROA study, supplemented with DSC, of caseins, synucleins and tau, together with several mutants of $\alpha$-synuclein and tau that cause neurodegenerative diseases. The ROA spectra of all these proteins are very similar to those of disordered poly(L-glutamic acid) at high pH and poly(L-lysine) at low pH [18,19]. Accordingly, the ROA spectra of disordered poly(L-lysine) and poly(L-glutamic acid) are reproduced in Fig. 1 to facilitate comparison with the protein ROA spectra. Largely on the basis of UVCD and VCD evidence, these two polypeptides are thought to contain substantial amounts of the PPII helical conformation, perhaps in the form of short
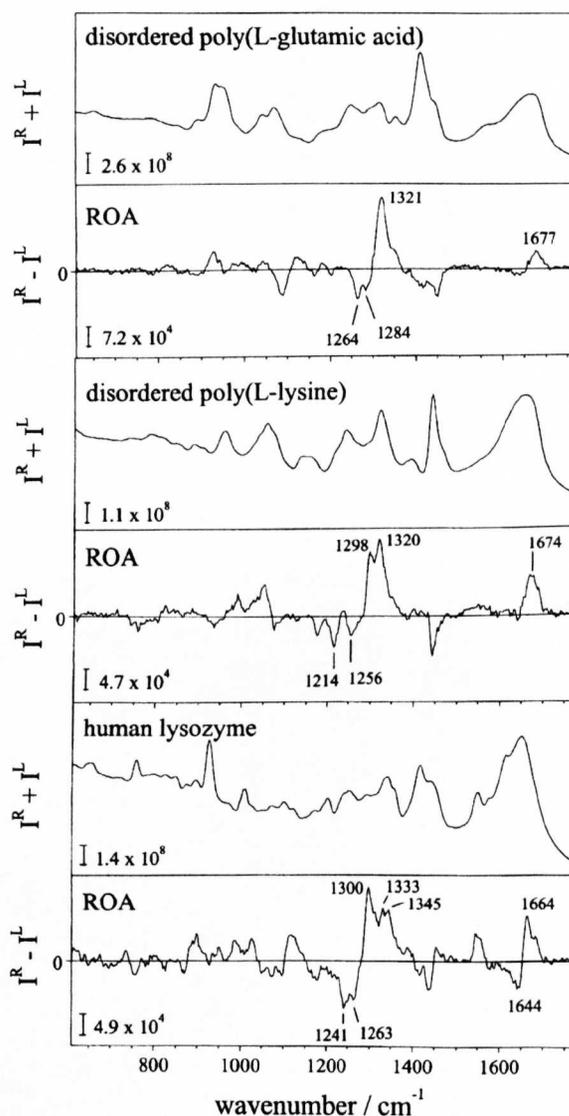


Fig. 1. The backscattered Raman and ROA spectra of disordered poly(L-glutamic acid) (top pair) and poly(L-lysine) (middle pair) in aqueous solution at pH 3.0 and 12.6, respectively, and of native human lysozyme at pH 5.4 (bottom pair).

segments interspersed with residues having other conformations [21,30–32]. We therefore consider their ROA spectra to show prominent bands characteristic of PPII structure, especially the strong positive ROA band at $\approx 1320$ cm$^{-1}$ [19]. Similar positive bands are observed in the ROA spectra of some β sheet proteins in the range $\approx 1315$–1325 cm$^{-1}$ that have been assigned [18] to PPII helical elements known from X-ray crystal structures to be present in some of the longer loops [33,34]. To date no reliable *ab initio* computations of the ROA spectrum of PPII helix have been performed, so our assignment of strong positive ROA at $\approx 1320$ cm$^{-1}$ to PPII structure relies mainly on the evidence outlined above.

In view of the close similarity of the ROA spectra of the casein, synuclein and tau proteins shown in this paper, we have reproduced in Fig. 1 from an earlier study [22], the ROA spectrum of a typical protein with a well-defined native fold, namely human lysozyme, in order to emphasize that different structural types of proteins usually give quite distinct ROA spectra. The ROA spectrum of human lysozyme contains many sharp bands characteristic of the different types of well-defined structural elements present. It is reassuring that there is no positive ROA band at $\approx 1320$ cm$^{-1}$ as the X-ray crystal structure contains no PPII helix [33]. However, such a band dominates the ROA spectrum of a destabilized intermediate of human lysozyme (produced on heating to 57 °C at pH 2.0) that forms prior to amyloid fibril formation and which prompted the suggestion, mentioned above, that PPII helix may be implicated in the generation of regular fibrils in amyloid disease [22].

## MATERIALS AND METHODS

### Materials

The β-casein was prepared from whole acid casein by the urea fractionation method of Aschaffenburg [35]. The κ-casein was prepared by adaptation of two other methods, each of which employs an acid precipitation stage to isolate the whole casein, a calcium precipitation stage to partially separate the Ca$^{2+}$-sensitive caseins from κ-casein, and an ethanol precipitation to isolate pure κ-casein. The method was essentially that of McKenzie & Wake [36] but instead of removing the excess Ca$^{2+}$ by precipitation with ammonium oxalate, the dialysis procedure of Talbot & Waugh [37] was employed, as this gives more control over ionic strength and a higher yield of the pure protein. Both proteins were shown to be better than 95% pure by alkaline urea PAGE, with only the β-casein showing slight contamination with a glycosylated form of κ-casein.

Recombinant wild-type human α-, β- and γ-synuclein, as well as the A30P and A53T mutants of α-synuclein, were purified to homogeneity, as described previously [38]. Proteins were prepared in a concentrated form by dialysis against 50 mM ammonium bicarbonate, followed by freeze-drying and reconstitution in the appropriate volume of water. Recombinant wild-type tau46 (corresponding to the 412-amino-acid isoform of human brain tau) and its P301L mutant were purified as described previously [15], except that the purified proteins were dialyzed against 25 mM Tris/HCl, pH 7.4, and further concentrated by Centricon (Millipore) filtration.

### Sample handling

The casein solutions were prepared at concentrations $\approx 50$ mg·mL$^{-1}$ in 50 mM phosphate buffer at pH 7.0 in small glass sample tubes, mixed with a little activated charcoal to remove traces of fluorescing impurities, and centrifuged. The solutions were subsequently filtered through 0.22 μm Millipore filters directly into quartz microfluorescence cells that were again centrifuged gently prior to mounting in the ROA instrument. Synuclein samples were prepared at $\approx 50$ mg·mL$^{-1}$ of protein in 50 mM Tris/HCl, pH 7.2. However, these solutions contained significant amounts of buffer salts due to their presence in the dry synuclein samples. Tau solutions were prepared at $\approx 30$ mg·mL$^{-1}$. Due to the smaller amounts of synuclein and tau available, treatment with charcoal was omitted and the solutions pipetted directly into the cells without microfiltration. Residual visible fluorescence from remaining traces of impurities, which can give large backgrounds in Raman spectra, was quenched by leaving the sample to equilibrate in the laser beam for several hours before acquiring ROA data.

The oligomeric state of the samples was not assessed at the high concentrations used for the ROA experiments and the possible effects of potential associations were not taken into account in the discussion of the results. This is justified from our experience that protein ROA spectra are generally insensitive to concentration, and even to oligomerization provided the intrinsic monomer conformations do not change, probably because ROA is sensitive mainly to local conformational features [18].

### ROA spectroscopy

The instrument used for the Raman and ROA measurements has a backscattering configuration, which is essential for aqueous solutions of biopolymers, and employs a single-grating spectrograph fitted with a backthinned CCD camera as detector and a holographic notch filter to block the Rayleigh line [39]. ROA is measured by synchronizing the Raman spectral acquisition with an electro-optic modulator, which switches the polarization of the incident argon-ion laser beam between right- and left-circular at a suitable rate. The spectra are displayed in analog-to-digital counter units as a function of Stokes wavenumber shift with respect to the exciting laser wavenumber. The ROA spectra are presented as raw circular intensity differences $I^R - I^L$ and the parent Raman spectra as raw circular intensity sums $I^R + I^L$, where $I^R$ and $I^L$ are the Raman-scattered intensities in right- and left-circularly polarized incident light, respectively. The experimental conditions for each measurement run were as follows: laser wavelength 514.5 nm; laser power at the sample $\approx 700$ mW; spectral resolution $\approx 10$ cm$^{-1}$; acquisition times $\approx 10$–20 h. The gaps in some of the synuclein ROA spectra arise from the removal of artefactual bands associated with intense polarized Raman bands from the significant amounts of buffer salts present.

### DSC measurements

The DSC measurements on β- and κ-casein were performed using a Microcal MCS calorimeter at the Hannah Research Institute: thermograms were recorded from 5 to 110 °C at a

scan rate of 1 °C·min$^{-1}$. The DSC measurements on the α-synuclein and tau proteins were performed using a Microcal MC2-D calorimeter by A. Cooper within the EPSRC/BBSRC funded facility at Glasgow University: thermograms were recorded from 15 to 100 °C at a scan rate of 1 °C·min$^{-1}$. The pH values were close to those used for the corresponding ROA measurements but the protein concentrations were much lower, ≈ 10 mg·mL$^{-1}$ for the Hannah instrument and ≈ 1 mg·mL$^{-1}$ for the Glasgow instrument (which is more sensitive). It was not possible to make DSC measurements on all of the proteins at the higher concentrations used for the ROA measurements due to the large amount of material required. However, sufficient quantities of β- and κ-casein were available, so as a check the measurements on these two proteins were repeated at ≈ 50 mg·mL$^{-1}$. The results were very similar to those obtained at the lower concentrations.

## RESULTS AND DISCUSSION

### ROA measurements on β- and κ-casein

The caseins constitute nearly 80% of bovine milk proteins. The major components, $\alpha_{S1}$-, $\alpha_{S2}$-, β- and κ-casein, occur in milk in the proportions (mass fractions) 0.37 : 0.09 : 0.41 : 0.13, respectively, as colloidal calcium phosphate micelles [40,41]. The monomers, which have molecular masses ≈ 19–25 kDa, are relatively unconstrained structures with very few disulfide links which are inter rather than intramolecular [42–44]. Early spectroscopic work suggested that caseins are largely 'structureless' with little extended secondary structure, but later UVCD studies suggested that, although largely 'random coil', $\alpha_{S1}$- and β-casein may contain ≈ 20% α helix and possibly a small amount of β sheet [45,46]. A conventional Raman study indicated ≈ 10% α helical structure and ≈ 20% β structure in both $\alpha_{S1}$- and β-casein, but different fine structure in the two Raman spectra suggested that their conformations are not identical [47]. UVCD and FTIR spectroscopy of κ-casein indicate ≈ 10–20% α helix and ≈ 30–40% β sheet structures with some evidence from UVCD and $^1$H-NMR studies on short peptides that the former is likely to be in the C-terminal half and the latter in the N-terminal half of the protein [29,48–51]. Sequence-based structure prediction methods suggest that the caseins are of the all β strand type, but that condensation into β sheets is inhibited by certain of the conserved features of the primary structure, allowing the proteins to retain an open and mobile rheomorphic conformation [3].

Here we report ROA measurements on β- and κ-casein. Although measurements were also attempted on $\alpha_{S1}$- and $\alpha_{S2}$-casein, these proteins had a tendency to aggregate in the laser beam, which prevented the acquisition of ROA data of sufficient quality for reliable analysis. A ROA spectrum of rather poor quality of an impure commercial sample of α-casein (composition undefined) was reported in an earlier study from which it was deduced that a large amount of PPII structure is present [19].

Figure 2 shows the room temperature backscattered Raman and ROA spectra of bovine β-casein (top pair) and κ-casein (bottom pair) at pH 7.0. Overall, the ROA spectra are very much alike, demonstrating that the basic structures of the proteins in aqueous solution are very similar. Both are
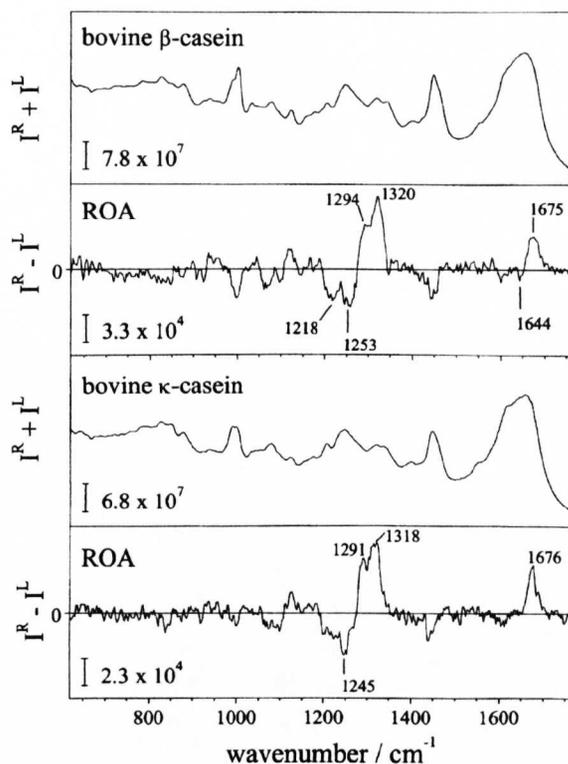


**Fig. 2. The backscattered Raman and ROA spectra of bovine β-casein (top pair) and κ-casein (bottom pair) in phosphate buffer, pH 7.0, measured at room temperature (≈ 20 °C).**

dominated by a strong positive ROA band centred at ≈ 1318–1320 cm$^{-1}$ in the extended amide III region, where normal vibrational modes containing largely $C_\alpha$–H and N–H deformations and the $C_\alpha$–N stretch usually contribute. A similar positive band at ≈ 1320 cm$^{-1}$ dominates the ROA spectra of disordered poly(L-glutamic acid) at pH 12.6 and poly(L-lysine) at pH 3.0 (Fig. 1). As these disordered polypeptides are thought to contain substantial amounts of the PPII helical conformation (see below), these β- and κ-casein ROA bands are therefore assigned to PPII structure.

The positive ROA bands in β- and κ-casein at ≈ 1290–1295 cm$^{-1}$ may originate in other types of loops and turns. A negative ROA band in the region ≈ 1238–1253 cm$^{-1}$ appears to be a reliable signature of β strand, individually or within β sheet, so the well-defined negative band at ≈ 1245 cm$^{-1}$ in the ROA spectrum of κ-casein is assigned here to β strand (rather than β sheet from the appearance of the amide I ROA, see below) [18]. The negative intensity in a similar region of the ROA spectrum of β-casein may have a similar origin. The two caseins also show significant negative ROA intensity at ≈ 1220 cm$^{-1}$ for which evidence is accumulating that this originates in a more hydrated form of β strand [18].

The positive bands at ≈ 1675 cm$^{-1}$ in the amide I region of the ROA spectra of β- and κ-casein, which originate mainly in the peptide C = O stretch, are characteristic of disordered structure, including the more static PPII type [18,19]. Regular β sheet is characterized by an amide I ROA couplet, negative at low wavenumber and positive at high and centred at ≈ 1655–1669 cm$^{-1}$ [18]. The absence of a clear negative

component here (although there is a hint) in the ROA spectra of β- and κ-casein may be evidence that, as suggested previously [3], the β-structure identified above mainly takes the form of unassociated β strands rather than β sheet.

These data suggest that the major conformational element present in β- and κ-casein is PPII helix. A significant amount of β strand may also be present, some of it hydrated, but little well-defined β sheet.

## ROA measurements on α-, β- and γ-synuclein

The α-, β- and γ-synucleins are related proteins of unknown function that range from 127 to 140 amino acids in length [9,52,53]. α-Synuclein is the major component of the filamentous lesions of Parkinson's disease, dementia with Lewy bodies and multiple system atrophy [25,26]. Synucleins lack cysteine or tryptophan residues. They have relatively unconstrained structures that are 'random coil' according to UVCD and other techniques [10–12]. Here we report ROA measurements on recombinant human versions of synucleins, together with the A30P and A53T mutants of α-synuclein that cause familial cases of Parkinson's disease. Unfortunately the quality of some of these synuclein ROA spectra is generally not as good as that of the caseins due in part to the high concentrations of buffer salts.

Figure 3 shows the backscattered Raman and ROA spectra of recombinant wild-type human α-synuclein (top pair) together with those of the A30P (middle pair) and A53T (bottom pair) mutants at pH 7.2. All three ROA spectra are very similar to each other, being dominated by a strong positive band centred at ≈ 1318–1320 cm$^{-1}$ assigned to PPII structure. They likewise have a single positive ROA band at ≈ 1675 cm$^{-1}$ in the amide I region assigned to disordered/PPII structure. Figure 4 shows the backscattered Raman and ROA spectra of β-synuclein (top pair) and γ-synuclein (bottom pair) at pH 7.2 that contain major features similar to those in the α-synucleins.

These data suggest that, as in the caseins, the major conformational element present in wild-type α-synuclein and the A30P and A53T mutants, as well as in β- and γ-synuclein, is PPII helix.

## ROA measurements on tau protein

Six isoforms of tau protein, ranging from 352 to 441 amino acids in length, are expressed in the adult human brain [54]. They fall into two classes, depending on the number of microtubule-binding repeats. Three isoforms have three repeats each and the other three isoforms have four repeats each. Depending on the isoforms, tau has either one (three-repeat forms) or two (four-repeat forms) cysteine residues. According to UVCD and other techniques, tau has a predominantly random coil structure with little or no α helix or β sheet [13–16]. Here we report ROA measurements on recombinant human four-repeat tau46 and its P301L mutant that causes frontotemporal dementia and Parkinsonism linked to chromosome 17 (FTDP-17). Tau46 corresponds to the 412-amino-acid isoform of human brain tau.

At neutral pH, the tau samples showed aggregation in the laser beam, with the aggregates falling to the bottom of the cell, so that the concentration of protein in solution decreased steadily with time. However, on reducing the pH to ≈ 4.3 no aggregation occurred, so that the ROA
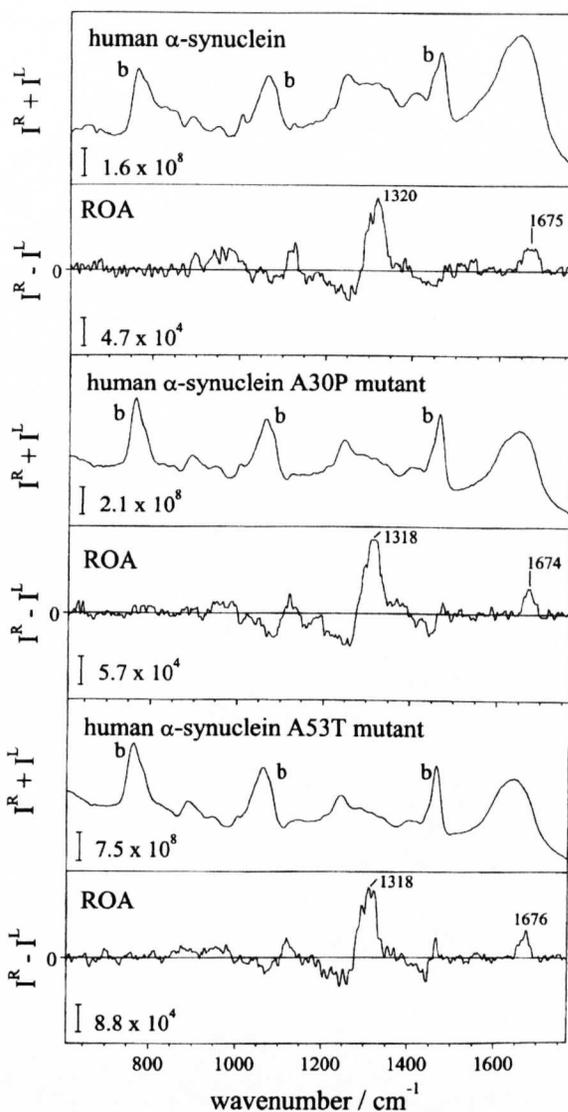


Fig. 3. **The backscattered Raman and ROA spectra of recombinant human wild-type α-synuclein (top pair), the A30P mutant (middle pair) and the A53T mutant (bottom pair) in Tris/HCl, pH 7.2, measured at room temperature.** The strong bands from buffer salts in the parent Raman spectra are marked with 'b'.

measurements were made at this reduced pH. As the native proteins are already in an unfolded state, such mild acidic conditions are unlikely to alter the conformation significantly. The backscattered Raman and ROA spectra of the wild-type and mutant tau46 are shown as the top and bottom pairs, respectively, in Fig. 5. Both ROA spectra show a strong positive ROA band centred at ≈ 1316–1318 cm$^{-1}$, indicating that a major conformational element is PPII helix like in the caseins and synucleins. They also show positive intensity in the range ≈ 1670–1675 cm$^{-1}$ characteristic of disordered/PPII structure. Some of the negative ROA intensity in the range ≈ 1240–1266 cm$^{-1}$ may be due to β strand.

These data suggest that, as in the caseins and synucleins, the major conformational element present in the wild-type
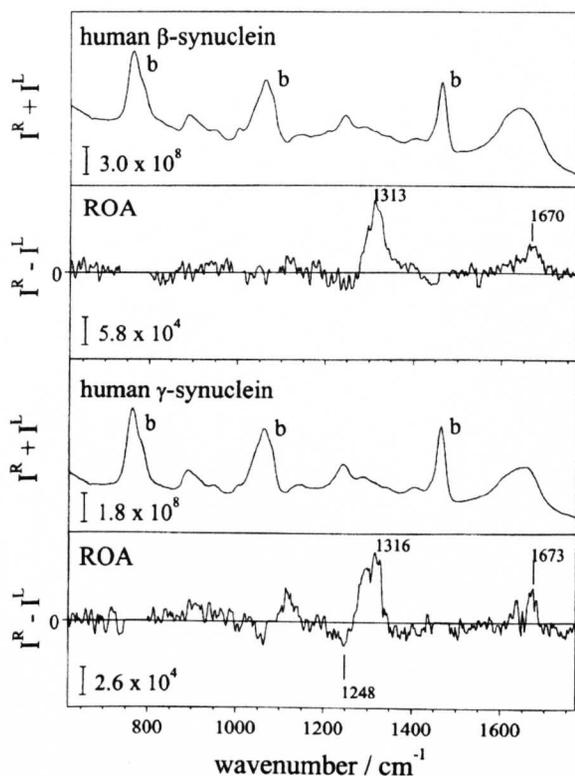
**Fig. 4. The backscattered Raman and ROA spectra of recombinant human β-synuclein (top pair) and γ-synuclein (bottom pair) in Tris/HCl, pH 7.2, measured at room temperature.** ROA data originating in artefacts from buffer bands have been cut out in some places.

**Fig. 5. The backscattered Raman and ROA spectra of recombinant human wild-type tau46 (top pair) and the tau 46 P301L mutant (bottom pair) in Tris/HCl with added HCl to reduce the pH to ≈ 4.3, measured at room temperature.**

and the P301L mutant of human tau 46 is PPII helix. Some β strand may also be present, but no β sheet.

## Caseins, synucleins and tau as rheomorphic proteins

The ROA data clearly show the caseins, synucleins and tau to have similar molecular structures which, from the presence of strong positive ROA bands in the range ≈ 1316–1320 cm$^{-1}$, may be based largely on the PPII helical conformation. There may also be some β strand in some of the proteins, especially β- and κ-casein judging by the well-defined negative ROA bands in these proteins in the range ≈ 1245 cm$^{-1}$, but little or no well-defined β sheet from the absence of a characteristic couplet in the amide I region. The caseins [46,55], synucleins [10] and tau [14] show no evidence of sharp denaturation to a more disordered structure on heating. We performed DSC measurements (data not shown) on β- and κ-casein, on wild-type α-synuclein, on the A30P and A53T mutants of α-synuclein, and on wild-type tau46. We found no evidence for a high-temperature thermal transition associated with cooperative unfolding. (In fact β-casein did show a weak concentration-dependent low-temperature thermal transition with a mid-point at ≈ 13 °C.)

These results indicate that the caseins, synucleins and tau are 'natively unfolded' structures in which the sequences are based largely on the PPII conformation and are held together in a loose noncooperative fashion. However, rather than describing them as 'random coil', the term 'rheomor-

phic' would seem to apply equally well to the synucleins and tau as it does to the caseins for which it was originally coined [3]. We attribute the lack of agreement between our present results and the earlier interpretations of the UVCD spectra of β- and κ-caseins (see above) to the fact that the basis sets of protein UVCD spectra used in the analysis do not normally include anything other than globular proteins with a well established X-ray crystal structure for which PPII structure is often not clearly distinguished from unordered structure. It can therefore not be relied upon to accurately represent the spectrum of a protein containing a large proportion of this conformation.

We envisage a rheomorphic protein to have the following general properties. The radius of gyration and hydrodynamic radius are ≈ two to four times larger than for a globular protein containing a similar number of residues, as observed in the caseins [3], synucleins [10], tau [14], prothymosin α [7] and the fibronectin-binding protein [8], and also in typical chemically denatured proteins [56–58]. Over extensive lengths of its sequence, the polypeptide chain is expected to be rather stiff, having a persistence length of ≈ 5–10 residues as reported for prothymosin α [7] and the fibronectin-binding protein [8]. In other parts of the molecule, there may be local interactions and small amounts of regular secondary structure but, as observed in some denatured proteins [59,60], interactions between remote parts of the sequence are expected to be minimal and many of the side chains are expected to have conformational flexibility. We do not consider the rheomorphic state of a

protein to be the same as the molten globule state as the latter is almost as compact as the folded state (radius of gyration and hydrodynamic radius ≈ 10–30% larger), has a hydrophobic core and contains a large amount of secondary structure [61,62].

Bowman–Birk protease inhibitors provide good examples of proteins which, despite having nonregular structures, are not natively unfolded. They are small single-chain proteins of molecular mass ≈ 7–9 kDa with seven disulfide links which stabilize a native fold comprising two tandem homologous domains [5]. Figure 6 shows the X-ray crystal structure (PDB code 1 pi2) of the soybean variant of this protein, together with its ROA spectrum measured earlier [19]. The general appearance of the ROA spectrum is quite similar to those of the caseins, synucleins and tau, except that it contains more detail as the fixed fold contains well-defined loops and turns plus a small amount of well-defined β sheet, together with fixed conformations for many of the side chains. As proteins belonging to different structural classes give quite different characteristic ROA band patterns [18], this suggests that the major conformational elements are similar and hence that the structures of the caseins, synucleins and tau may be envisaged as more open, hydrated, longer-chain (and nonglobular) versions of the structure of the Bowman–Birk inhibitor in Fig. 6. The X-ray crystal structure 1 pi2 reveals that the φ,ψ angles of most of the residues of the Bowman–Birk inhibitor are distributed fairly evenly over the β- and PPII-regions of the
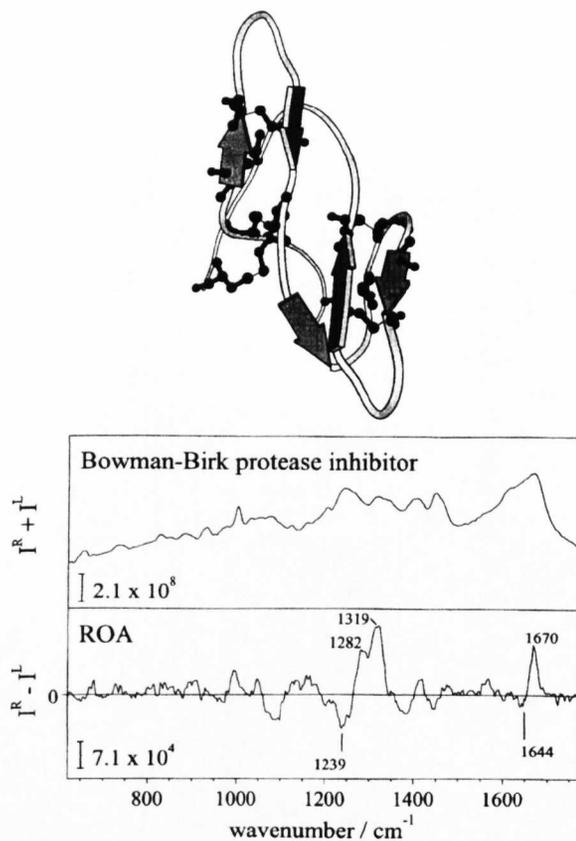


Fig. 6. A MOLSCRIPT diagram [67] of the X-ray crystal structure of soybean Bowman–Birk inhibitor (PDB code 1 pi2) together with its backscattered Raman and ROA spectra in acetate buffer, pH 5.4.

Ramachandran surface, so the same may be true for the constituent residues of the caseins, synucleins and tau.

## Relative propensities for β-fibril formation

It has been suggested recently that, as it is extended, flexible, lacks intrachain hydrogen bonds and is fully hydrated in aqueous solution, PPII helix may be the 'killer conformation' in amyloid diseases [22]. This is because elimination of water molecules between extended polypeptide chains with fully hydrated C = O and N–H groups to form β sheet hydrogen bonds is a highly favourable process entropically, and as strands of PPII helix are close in conformation to β strands, they would be expected to readily undergo this type of aggregation with each other and also with the edges of established β sheet. The more dynamic type of disorder associated with the true random coil is expected to lead to amorphous aggregates rather than ordered fibrils, as is observed in most examples of protein aggregation. However, although the presence of significant amounts of PPII structure may be necessary for the formation of regular fibrils, other factors must be important as, of all the rheomorphic proteins studied here, only α-synuclein is known to readily form typical amyloid cross β fibrils [11,12,63]. (The presence or otherwise of β sheet, and hence of a cross β substructure, in filamentous aggregates of tau remains unclear [14,64].)

For example, Biere *et al.* [12] suggested that the failure of β-synuclein to fibrillize under their conditions could be due to its lack of a sequence present in α-synuclein (residues 72–84) which, according to structure prediction methods, has a high β sheet forming propensity. And Holt & Sawyer [3] suggested that the abundance of glutamine residues in the β-caseins may act to prevent β sheet formation by competitive side-chain–backbone hydrogen bonding interactions, thus helping to maintain, along with the abundance of proline residues, the open conformation of the protein. The finding that a combination of low mean hydrophobicity and high net charge are important prerequisites for proteins to remain natively unfolded [1] may be especially pertinent here. One possible example of the significance of charge is the observation that removal of the highly charged anionic C-terminal region from α-synuclein results in more rapid fibril formation than for the wild-type and the A53T and A30P mutants [11,38]. Another is the increased fibrillogenicity of mouse α-synuclein compared with human that may be due in part to the decreased charge and polarity in the C-terminal region due to a difference of five residues in this region [65].

Vigorous shaking is required to induce rapid amyloid fibril formation from full-length α-synuclein [11]. Shaking may lead to the shearing of α-synuclein assemblies, which then function as seeds, resulting in a marked acceleration of filament formation. On the other hand, Serio *et al.* [66] found that only modest rotation of the yeast prion protein Sup35 was effective in inducing amyloid fibril formation. These observations could be consistent with the presence of large amounts of PPII structure, as any agitation which produces fluid flow, as in a circular motion, would tend to align the PPII helical sequences, thereby making it more favourable for them to aggregate into ordered β sheet. These two possible mechanisms (generation of new seeds plus alignment of PPII sequences) could strongly reinforce each other.

## CONCLUSIONS

This study has shown that the casein milk proteins, the brain proteins synuclein and tau, as well as mutants of α-synuclein and tau associated with inherited forms of neurodegenerative disease, all have a very similar type of structure, possibly based on the PPII conformation, and which may be envisaged as a more open version of the X-ray crystal structure of the Bowman–Birk inhibitor. The rheomorphic character imparted by large amounts of extended, flexible, hydrated PPII sequences may be important for the function of these proteins. Although disorder of the PPII type may be an essential requirement for the formation of regular fibrils [22], our results suggest that the presence of a large amount of PPII structure does not necessarily impart a fibrillogenic character, as neither full-length caseins, nor β- and γ-synuclein, show a significant propensity for amyloid fibril formation. Further understanding of fibrillogenic propensity should therefore be sought not so much in conformational differences but in the various properties of residues and how these modulate the association characteristics of particular sequences.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Uversky, V.N., Gillespie, J.R. & Fink, A.L. (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* 41, 415–427.
2. Wright, P.E. & Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* 293, 321–331.
3. Holt, C. & Sawyer, L. (1993) Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the α$_{S1}$-, β- and κ-caseins. *J. Chem. Soc. Faraday Trans.* 89, 2683–2692.
4. Holt, C. & van Kemenade, M.J.J.M. (1989) Interaction of phosphoproteins with calcium phosphates. In *Calcified Tissue* (Hukins, D.W.L., ed.), pp. 175–213. Macmillan, London, UK.
5. de la Sierra, I.L., Quillien, L., Flecker, P., Gueguen, J. & Brunie, S. (1999) Dimeric crystal structure of a Bowman–Birk protease inhibitor from pea seeds. *J. Mol. Biol.* 285, 1195–1207.
6. Braun, W., Vašák, M., Robbins, A.H., Stout, C.D., Wagner, G., Kägi, J.H.R. & Wüthrich, K. (1992) Comparison of the NMR solution structure and the X-ray crystal structure of rat metallothionein-2. *Proc. Natl Acad. Sci. USA* 89, 10124–10128.
7. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H.R., Müller-Frohne, M., Zirwer, D., Czarnecki, J. & Damaschun, G. (1995) Prothymosin α: a biologically active protein with a random coil conformation. *Biochemistry* 34, 13211–13218.
8. Penkett, C.J., Redfield, C., Jones, J.A., Dodd, I., Hubbard, J., Smith, R.A.G., Smith, L.J. & Dobson, C.M. (1998) Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. *Biochemistry* 37, 17054–17067.
9. Jakes, R., Spillantini, M.G. & Goedert, M. (1994) Identification of two distinct synucleins from human brain. *FEBS Lett.* 345, 27–32.
10. Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A. & Lansbury, P.T. Jr (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35, 13709–13715.
11. Serpell, L.C., Berriman, J., Jakes, R., Goedert, M. & Crowther, R.A. (2000) Fiber diffraction of synthetic α-synuclein filaments shows amyloid-like cross-β conformation. *Proc. Natl. Acad. Sci. USA* 97, 4897–4902.
12. Biere, A.L., Wood, S.J., Wypych, J., Steavenson, S., Jiang, Y., Anafi, D., Jacobsen, F.W., Jarosinski, M.A., Wu, G.M., Louis, J.C., Martin, F., Narhi, L.O. & Citron, M. (2000) Parkinson's disease-associated α-synuclein is more fibrillogenic than β- and γ-synuclein and cannot cross-seed its homologs. *J. Biol. Chem.* 275, 34574–34579.
13. Cleveland, D.W., Hwo, S.H. & Kirschner, M.W. (1977) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *J. Mol. Biol.* 116, 227–247.
14. Schweers, O., Schönbrunn-Hanebeck, E., Marx, A. & Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for β-structure. *J. Biol. Chem.* 269, 24290–24297.
15. Goedert, M., Jakes, R. & Crowther, R.A. (1999) Effects of frontotemporal dementia FTDP-17 mutations on heparin-induced assembly of tau filaments. *FEBS Lett.* 450, 306–311.
16. Barghorn, S., Zheng-Fischhöfer, Q., Ackmann, M., Biernat, J., von Bergen, M., Mandelkow, E.M. & Mandelkow, E. (2000) Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias. *Biochemistry* 39, 11714–11721.
17. Poland, D. & Scheraga, H.A. (1970) *Theory of Helix-Coil Transitions in Biological Macromolecules*. Academic Press, New York.
18. Barron, L.D., Hecht, L., Blanch, E.W. & Bell, A.F. (2000) Solution structure and dynamics of biomolecules from Raman optical activity. *Prog. Biophys. Mol. Biol.* 73, 1–49.
19. Smyth, E., Syme, C.D., Blanch, E.W., Hecht, L., Vašák, M. & Barron, L.D. (2000) Solution structure of native proteins with irregular folds from Raman optical activity. *Biopolymers* 58, 138–151.
20. Sreerama, N. & Woody, R.W. (2000) Circular dichroism of peptides and proteins. In *Circular Dichroism. Principles and Applications* (Berova, N., Nakanishi, K. & Woody, R.W., eds), pp. 601–620. Wiley, New York.
21. Keiderling, T.A. (2000) Peptide and protein conformational studies with vibrational circular dichroism and related spectroscopies. In *Circular Dichroism. Principles and Applications* (Berova, N., Nakanishi, K. & Woody, R.W., eds), pp. 621–666. Wiley, New York.
22. Blanch, E.W., Morozova-Roche, L.A., Cochran, D.A.E., Doig, A.J., Hecht, L. & Barron, L.D. (2000) Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. *J. Mol. Biol.* 301, 553–563.
23. Holt, C., Wahlgren, N.M. & Drakenberg, T. (1996) Ability of a β-casein phosphopeptide to modulate the precipitation of calcium phosphate by forming amorphous dicalcium phosphate nanoclusters. *Biochem. J.* 314, 1035–1039.
24. Holt, C., Timmins, P.A., Errington, N. & Lever, J. (1998) A core-shell model of calcium phosphate nanoclusters derived from sedimentation equilibrium and small angle X-ray and neutron scattering measurements. *Eur. J. Biochem.* 252, 73–78.
25. Spillantini, M.G., Schmidt, M.L., Lee, V.M.Y., Trojanowski, J.Q., Jakes, R. & Goedert, M. (1997) α-Synuclein in Lewy bodies. *Nature* 388, 839–840.
26. Goedert, M. (1999) Filamentous nerve cell inclusions in neurodegenerative diseases: tauopathies and α-synucleinopathies. *Phil. Trans. Roy. Soc. Lond. B* 354, 1101–1118.
27. Hirokawa, N. (1994) Microtubule organization and dynamics dependent on microtubule-associated proteins. *Curr. Op. Cell Biol.* 6, 74–81.

28. Niewold, T.A., Murphy, C.L., Hulskamp-Koch, C.A.M., Tooten, C.J. & Gruys, E. (1999) Casein related amyloid, characterisation of a new and unique amyloid protein isolated from bovine corpora amylacea. *Int. J. Exp. Clin. Invest.* 6, 244–249.

29. Farrell, H.M. Jr, Qi, P.X., Brown, E.M., Cooke, P.H., Tunick, M.H., Wickham, E.D. & Unruth, J.J. Molten globule structures in milk proteins: implications for potential new structure-function relationships. J. Dairy Sci., in press.

30. Tiffany, M.L. & Krimm, S. (1968) New chain conformations of poly (glutamic acid) and polylysine. *Biopolymers* 6, 1379–1382.

31. Woody, R.W. (1992) Circular dichroism and conformations of unordered polypeptides. *Adv. Biophys. Chem.* 2, 37–39.

32. Keiderling, T.A., Silva, R.A.G.D., Yoder, G. & Dukor, R.K. (1999) Vibrational circular dichroism of selected oligopeptide conformations. *Bioorg. Med. Chem.* 7, 133–141.

33. Adzhubei, A.A. & Sternberg, M.J.E. (1993) Left-handed poly-proline II helices commonly occur in globular proteins. *J. Mol. Biol.* 229, 472–493.

34. Stapley, B.J. & Creamer, T.P. (1999) A survey of left-handed polyproline II helices. *Protein Sci.* 8, 587–595.

35. Aschaffenburg, R. (1963) Preparation of β-casein by a modified urea fractionation method. *J. Dairy Res.* 30, 259–260.

36. McKenzie, H.A. & Wake, R.G. (1961) An improved method for the isolation of κ-casein. *Biochim. Biophys. Acta* 47, 240–242.

37. Talbot, B. & Waugh, D.F. (1970) Micellar forming characteristics of monomeric and covalent polymeric κ-caseins. *Biochemistry* 9, 2807–2813.

38. Crowther, R.A., Jakes, R., Spillantini, M.G. & Goedert, M. (1998) Synthetic filaments assembled from C-terminally truncated α-synuclein. *FEBS Lett.* 436, 309–312.

39. Hecht, L., Barron, L.D., Blanch, E.W., Bell, A.F. & Day, L.A. (1999) Raman optical activity instrument for studies of biopolymer structure and function. *J. Raman Spectrosc.* 30, 815–825.

40. Holt, C. (1992) Structure and stability of the bovine casein micelle. *Adv. Protein. Chem.* 43, 63–151.

41. Rollema, H.S. (1992) Casein association and micelle formation. In *Advanced Dairy Chemistry-I: Proteins* (Fox, P.F., ed.), pp. 111–140. Elsevier Applied Science, London.

42. Groves, M.L., Dower, H.J. & Farrell, H.M. Jr (1992) Reexaination of the polymeric distribution of κ-casein isolated from bovine milk. *J. Prot. Chem.* 11, 21–28.

43. Rasmussen, L.K., Højrup, P. & Petersen, T.E. (1992) Localisation of two interchain disulphide bridges in dimers of bovine α$_{S2}$-casein. Parallel and antiparallel alignments of the polypeptide chains. *Eur. J. Biochem.* 203, 381–386.

44. Rasmussen, L.K., Højrup, P. & Petersen, T.E. (1992) The multimeric structure and disulphide bonding pattern of bovine κ-casein. *Eur. J. Biochem.* 207, 215–222.

45. Creamer, L.K., Richardson, T. & Parry, D.A.D. (1981) Secondary structure of bovine α$_{S1}$- and β-casein in solution. *Arch. Bioch. Biophys.* 211, 689–696.

46. Graham, E.R.B., Malcolm, G.N. & McKenzie, H.A. (1984) On the isolation and conformation of bovine β-casein. *Int. J. Biol. Macromol.* 6, 155–161.

47. Byler, D.M. & Susi, H. (1988) Application of computerized infrared and Raman spectroscopy to conformation studies of casein and other food proteins. *J. Indust. Microbiol.* 3, 73–88.

48. Raap, J., Kerling, K.E.T., Vreeman, H. & Visser, S. (1983) Peptide substrates for chymosin (rennin): conformational studies of κ-casein and some κ-casein-related oligopeptides by circular dichroism and secondary structure prediction. *Arch. Biochem. Biophys.* 221, 117–124.

49. Griffin, M.C.A., Price, J.C. & Martin, S.R. (1986) Effect of alcohols on the structures of caseins: circular dichroism studies of κ-casein. *Int. J. Biol. Macromol.* 8, 367–371.

50. Ono, T., Yada, R., Yutani, K. & Nakai, S. (1987) Comparison of conformations of κ-casein, para-κ-casein and glycomacropeptide. *Biochim. Biophys. Acta* 911, 318–325.

51. Creamer, L.K., Plowman, J.E., Liddell, M.J., Smith, M. & Hill, J.P. (1998) Micelle stability: κ-casein structure and function. *J. Dairy Sci.* 81, 3004–3012.

52. Uéda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., Otero, D.A.C., Kondo, J., Ihara, Y. & Saitoh, T. (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl Acad. Sci. USA* 90, 11282–11286.

53. Ji, H., Liu, Y.E., Jia, T., Wang, M., Liu, J., Xiao, G., Joseph, B.K., Rosen, C. & Shi, Y.E. (1998) Identification of a breast cancer-specific gene, *BCSG1*, by direct differential cDNA sequencing. *Cancer Res.* 57, 759–764.

54. Goedert, M., Spillantini, M.G., Jakes, R., Rutherford, D. & Crowther, R.A. (1989) Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibril-lary tangles of Alzheimer's disease. *Neuron* 3, 519–526.

55. Paulsson, M. & Dejmek, P. (1990) Thermal denaturation of whey proteins in mixtures with caseins studied by differential scanning calorimetry. *J. Dairy Sci.* 73, 590–600.

56. Petrescu, A.J., Receveur, V., Calmettes, P., Durand, D. & Smith, J.C. (1998) Excluded volume in the configurational distribution of a strongly-denatured protein. *Protein Sci.* 7, 1396–1403.

57. Russo, D., Durand, D., Desmadril, M. & Calmettes, P. (2000) Study of thermally and chemically unfolded conformations of a small beta-protein by means of small-angle neutron scattering. *Physica* B276, 520–521.

58. Damaschun, H., Gast, K., Hahn, U., Krober, R., Müller-Frohne, M., Zirwer, D. & Damaschun, G. (1997) Conformation of thermally denatured Rnase T1 with intact disulphide bonds: a study by small-angle X-ray scattering. *Biochim. Biophys. Acta* 1340, 235–244.

59. Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J. & Dobson, C.M. (1997) Structural and dynamic properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8M urea. *Biochemistry* 36, 8977–8991.

60. Gillespie, J.R. & Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* 268, 158–169.

61. Ptitsyn, O.B. (1995) Molten globule and protein folding. *Adv. Prot. Chem.* 47, 83–229.

62. Arai, M. & Kuwajima, K. (2000) The role of the molten globule state in protein folding. *Adv. Prot. Chem.* 53, 209–282.

63. Conway, K.A., Harper, J.D. & Lansbury, P.T. (2000) Fibrils formed in vitro from α-synuclein and two mutant forms linked to Parkinson's disease are typical amyloid. *Biochemistry* 39, 2552–2563.

64. Von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E.M. & Mandelkow, E. (2000) Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ([306]VQIVYK[311]) forming β structure. *Proc. Natl Acad. Sci. USA* 97, 5129–5134.

65. Rochet, J.-C., Conway, K.A. & Lansbury, P.T. Jr (2000) Inhibition of fibrillization and accumulation of prefibrillar oligomers in mixtures of human and mouse α-synuclein. *Biochemistry* 39, 10619–10626.

66. Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F. & Lindquist, S.L. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* 289, 1317–1321.

67. Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946–950.

# Solution structures of potato virus X and narcissus mosaic virus from Raman optical activity

Ewan W. Blanch,[1] David J. Robinson,[2] Lutz Hecht,[1] Christopher D. Syme,[1] Kurt Nielsen[3] and Laurence D. Barron[1]

[1] Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

[2] Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

[3] Department of Chemistry, DTU 207, Technical University of Denmark, DK-2800 Lyngby, Denmark

Potato virus X (PVX) and narcissus mosaic virus (NMV) were studied using vibrational Raman optical activity (ROA) in order to obtain new information on the structures of their coat protein subunits. The ROA spectra of the two intact virions are very similar to each other and similar to that of tobacco mosaic virus (TMV) studied previously, being dominated by signals characteristic of proteins with helix bundle folds. In particular, PVX and NMV show strong positive ROA bands at $\sim 1340$ cm$^{-1}$ assigned to hydrated $\alpha$-helix and perhaps originating in surface exposed helical residues, together with less strong positive ROA intensity in the range $\sim 1297$–$1312$ cm$^{-1}$ assigned to $\alpha$-helix in a more hydrophobic environment and perhaps originating in residues at helix–helix interfaces. The positive $\sim 1340$ cm$^{-1}$ ROA band of TMV is less intense than those of PVX and NMV, suggesting that TMV contains less hydrated $\alpha$-helix. Small differences in other spectral regions reflect differences in some loop, turn and side-chain compositions and conformations among the three viruses. A pattern recognition program based on principal component analysis of ROA spectra indicates that the coat protein subunit folds of PVX and NMV may be very similar to each other and similar to that of TMV. These results suggest that PVX and NMV may have coat protein subunit structures based on folds similar to the TMV helix bundle and hence that the helical architecture of the PVX and NMV particles may be similar to that of TMV but with different structural parameters.

## Introduction

Potato virus X (PVX) and narcissus mosaic virus (NMV) belong to the *Potexvirus* genus of helical plant viruses, PVX being the type species (Brunt *et al.*, 2000). They contain genomes of single-stranded positive-sense RNA encapsidated within flexuous filamentous particles, $\sim 500$ nm long and $\sim 13$–14 nm in diameter, with helical symmetry and a pitch of $\sim 3\cdot4$ nm. In comparison, the *Tobamovirus* genus of helical plant viruses, to which tobacco mosaic virus (TMV) belongs, also contain single-stranded positive-sense RNA but the particles are rigid rods $\sim 300$ nm long and $\sim 18$ nm in diameter with a helical pitch of $\sim 2\cdot3$ nm. Although detailed information about the genome organization, the amino acid sequences of the encoded proteins and the sequence homology between members of the *Potexvirus* genus is available [Brunt *et al.* (2000)

and references therein], little is known about the conformation of the coat protein subunits and their packing arrangements in the virus particles. Unlike the *Tobamoviruses*, there are no X-ray diffraction data of sufficient resolution to show molecular conformational detail, but ultraviolet circular dichroism (UVCD) measurements suggest $\sim 45\%$ $\alpha$-helix content and $\sim 5\%$ $\beta$-sheet for both PVX (Homer & Goodman, 1975) and NMV (Wilson *et al.*, 1991). A model of the PVX coat protein tertiary structure has been suggested based on the predicted secondary structure content, the principles of secondary structure packing, and the known biochemical, immunological and tritium bombardment data (Baratova *et al.*, 1992a).

The lack of information about the molecular structure of potexviruses highlights the lack of physical techniques applicable to the determination of virus structure and function at the molecular level. A novel spectroscopic technique called Raman optical activity (ROA) appears to be promising for such studies and has recently been applied to intact viruses; specifically filamentous bacteriophages (Blanch *et al.*, 1999),

**Author for correspondence:** Laurence Barron.

Fax: +44 141 330 4888. e-mail laurence@chem.gla.ac.uk

and tobacco rattle virus (TRV) and tobacco mosaic virus (TMV) (Blanch *et al.*, 2001a). Here we report an ROA study of the DX strain of PVX (Jones, 1982) and the MC strain of NMV (Wilson *et al.*, 1991) from which new information about the folds of the coat proteins is deduced, mainly from a comparison with the ROA spectrum of the U1 strain of TMV.

## A brief review of Raman spectroscopy and ROA

Since the ROA technique is unfamiliar to most virologists, a brief description may be helpful. Raman spectroscopy itself is a form of vibrational spectroscopy, complementary to infrared absorption, which provides vibrational spectra of molecules by means of inelastic scattering of visible or ultraviolet laser light. During the so-called Stokes Raman scattering event, the interaction of the molecule with the incident laser photon of energy $\omega$, where $\omega$ is its angular frequency, can leave the molecule in an excited vibrational state of energy $\hbar\omega_v$, with a corresponding energy loss, and hence a shift to lower angular frequency $\omega - \omega_v$, of the scattered photon. Therefore, by analysing the spectrum of the scattered light with a visible or ultraviolet spectrometer, a complete vibrational spectrum of the molecule may be obtained. Conventional Raman spectroscopy has a number of favourable characteristics, including the fact that water is an excellent solvent for Raman studies, which have led to many applications in biochemistry (Carey, 1982; Miura & Thomas, 1995). It has proved especially valuable in structural virology on account of its ability to provide information about both protein and nucleic acid constituents of intact virions (Thomas, 1987, 1999).

ROA is a novel form of Raman spectroscopy that is sensitive to chirality, meaning handedness, in molecular structure. The two distinguishable mirror-image forms of a chiral molecule are called enantiomers. Biomolecules such as proteins and nucleic acids are chiral on account of the intrinsic chirality of their constituent amino acids (almost exclusively the L-enantiomers) and sugars (almost exclusively the D-enantiomers), respectively. Raman spectroscopy becomes sensitive to chirality by utilizing circularly polarized light: thus ROA measures small differences in the Raman spectra of chiral molecules acquired using right- and left-circularly polarized incident laser light. ROA bears the same relation to conventional Raman spectroscopy as does the widely used biochemical technique of UVCD to conventional visible and ultraviolet absorption spectroscopy. The advantage of ROA is that it provides *vibrational* optical activity spectra which contain much more stereochemical information than the *electronic* optical activity spectra provided by UVCD (just as conventional vibrational spectra, Raman scattering or infrared absorption, contain much more information about molecular structure than conventional visible and ultraviolet absorption spectra). Recent general reviews of ROA include Nafie (1997) and Barron & Hecht (2000).

ROA has recently been applied to biomolecules such as proteins, carbohydrates, nucleic acids and viruses. This work

has been reviewed by Barron *et al.* (2000). On account of its sensitivity to chirality, ROA is proving to be an incisive probe of biomolecular structure and dynamics. One reason for this is that the largest ROA signals are often associated with the most rigid chiral parts of the structure. In proteins these are usually within the peptide backbone and often give rise to ROA band patterns characteristic of the peptide backbone conformation, unlike the parent Raman spectra which are often dominated by bands from the amino acid side-chains, which often obscure the peptide backbone bands. As well as bands arising from secondary structure, protein ROA spectra also contain bands from loops and turns and so can provide information about the tertiary fold. A few ROA bands from side-chains also appear, including one for tryptophan from which the absolute stereochemistry can be deduced, even in the coat proteins of intact viruses (Blanch *et al.*, 2001b). Carbohydrate ROA spectra provide information about anomeric preference, the pattern of OH substituents and the nature of the glycosidic link, and can detect extended secondary structure in polysaccharides. Nucleic acid ROA spectra provide information about base stacking, the mutual orientations of sugar and base rings, and the sugar–phosphate backbone conformations.

## Methods

PVX, strain DX, was propagated in *Nicotiana tabacum* Samsun NN, and purified as described by Abou Haidar *et al.* (1998). NMV, strain MC, was propagated in *Chenopodium quinoa*, and purified as described by Goodman (1973) for PVX, followed by sedimentation in a 10–40 % sucrose gradient at 20000 r.p.m. for 140 min in a Beckman SW28 rotor.

Both viruses were studied as solutions at $\sim$ 25 mg/ml in 15 mM phosphate buffer at pH 7·4 held in small quartz microfluorescence cells at ambient temperature ($\sim$ 20 °C). The ROA measurements were performed using an instrument described previously (Hecht *et al.*, 1999). It has a backscattering configuration and employs a single-grating spectrograph fitted with a backthinned charge coupled device (CCD) camera as detector and an edge filter to block the Rayleigh line. The small ROA signals are accumulated by synchronizing the Raman spectral acquisition with an electrooptic modulator that switches the polarization of the argon-ion laser beam between right- and left-circular at a suitable rate. The spectra are displayed in analogue-to-digital counter units as a function of the Stokes Raman wavenumber shift with respect to the exciting laser wavelength. The ROA spectra, which have undergone minimal smoothing (2 point FFT), are presented as circular intensity differences $I^R - I^L$ and the parent Raman spectra as circular intensity sums $I^R + I^L$, where $I^R$ and $I^L$ are the Raman-scattered intensities in right- and left-circularly polarized incident light, respectively. The experimental conditions were as follows: laser wavelength 514·5 nm; laser power at the sample $\sim$ 700 mW; spectral resolution $\sim$ 10 cm$^{-1}$; acquisition time $\sim$ 48 h.

## Results and Discussion

### The ROA spectral details and their structural assignments

The backscattered Raman and ROA spectra of PVX and NMV are shown as the top and middle pairs of spectra,
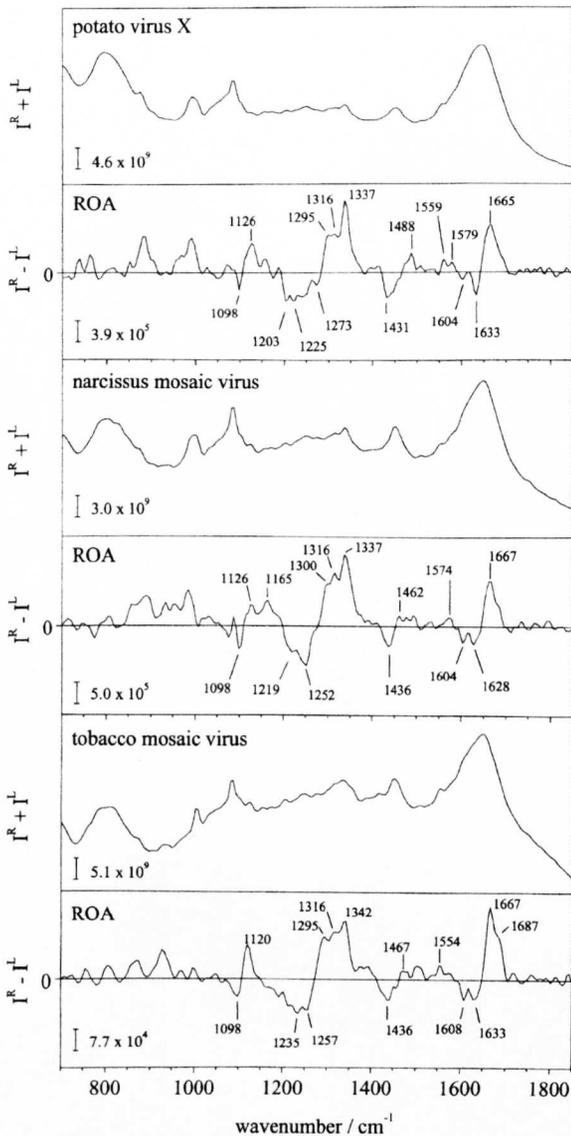
**Fig. 1.** Backscattered Raman ($I^R + I^L$) and ROA ($I^R - I^L$) spectra of PVX (top pair), NMV (middle pair) and TMV (bottom pair) in phosphate buffer at pH 7·4.

respectively, in Fig. 1. An account of most of the ROA assignments given in what follows can be found in Barron *et al.* (2000). The ROA spectra of the two viruses are dominated by protein bands. The general appearance of the ROA spectra of both viruses suggests that both contain a large amount of α-helix. In particular, both show an ROA couplet, negative at ∼ 1630 and positive at ∼ 1666 cm⁻¹, originating in peptide amide I vibrations involving mainly C=O stretching that is characteristic of α-helix. Positive protein ROA bands in the range ∼ 1295–1312 cm⁻¹ together with positive bands in the range ∼ 1337–1345 cm⁻¹, originating in extended amide III vibrations involving various contributions from peptide back-

bone N–H and $C_\alpha$–H deformations and the $C_\alpha$–N stretch, are also characteristic of α-helix. The former appear to arise from α-helix in a hydrophobic environment and the latter to a hydrated form of α-helix (Blanch *et al.*, 1999). Both PVX and NMV show strong positive ROA bands at ∼ 1337 cm⁻¹ together with weaker positive intensity centred ∼ 1295–1300 cm⁻¹ which are therefore assigned to hydrated and hydrophobic α-helix, respectively. Such an ROA band pattern is characteristic of helix bundle proteins in which most residues are surface exposed and hydrated and a smaller number are within hydrophobic interfaces between helices. One difference between the ROA spectra of the two viruses is the presence of a significant negative band at ∼ 1252 cm⁻¹ in the ROA spectrum of NMV assigned to β-strand, whereas PVX shows rather weaker less well-defined negative ROA here. However, both viruses show negative ROA bands at ∼ 1220 cm⁻¹ assigned to a hydrated form of β-strand. Differences in various weak bands reflect differences in the details of loop and turn sequences.

ROA bands from the RNA are expected to be much weaker than from the protein due to the low RNA contents (∼ 6% of the total particle mass in each case). It is possible that the weak positive bands at ∼ 1488 and 1579 cm⁻¹ in the ROA spectrum of PVX arise from the nucleic acid bases since RNA and DNA often show large ROA in these regions (Bell *et al.*, 1997, 1998) but proteins do not. Hints of similar bands appear in the ROA spectrum of NMV. More specifically, these bands may arise from guanine and adenine since bands at similar wavenumbers assigned to these bases in the viral nucleic acid are seen in the conventional transparent and resonance Raman spectra of intact viruses (Thomas, 1999), and naturally occurring nucleic acids show strong positive ROA at ∼ 1481 cm⁻¹ also assigned to guanine and adenine (Bell *et al.*, 1998).

The protein ROA data are consistent with the UVCD results, which predicted large amounts of α-helix with little β-sheet (Homer & Goodman, 1975; Wilson *et al.*, 1991). Furthermore, the overall appearances of the ROA spectra of PVX and NMV are very similar to each other and to that of TMV reported recently (Blanch *et al.*, 2001a) for which the coat protein subunits are based on four-helix bundles containing both water-exposed residues and residues at hydrophobic helix–helix interfaces (Namba *et al.*, 1989; Stubbs, 1999). One difference, however, is that the positive ∼ 1337 cm⁻¹ ROA bands in PVX and NMV are significantly more intense than the corresponding band in TMV, which suggests that PVX and NMV contain more hydrated α-helix than TMV. For convenience, a re-measured ROA spectrum of TMV (strain U1) of better quality than that recorded in the earlier study is shown at the bottom of Fig. 1. Our results therefore suggest that the coat protein folds of PVX and NMV are similar to that of TMV, being based on a helix bundle, but with differences of detail resulting from differences in the appended loops and turns and from the extra sequences present in PVX and NMV (the coat proteins of the DX strain of PVX, the MC strain of

NMV and the U1 strain of TMV contain 237, 240 and 159 amino acids, respectively). Our results do not support the model of Baratova *et al.* (1992a) for the PVX coat proteins, which may contain too much well-defined $\beta$-sheet (which is predicted at both the N- and C-terminal ends). However, unlike the N-terminal regions of PVX which tritium planigraphy shows are exposed, the C-terminal regions of PVX are virtually inaccessible to tritium labelling in the intact virion (Baratova *et al.*, 1992b), but become accessible after $\sim$ 20 N-terminal amino acids are proteolyticaly removed (Baratova *et al.*, 1992a). For TMV, on the other hand, tritium planigraphy shows that both N- and C-terminal regions are exposed (Goldanskii *et al.*, 1988), as expected from the X-ray structure (Stubbs, 1999). Some of the additional $\sim$ 80 residues in PVX and NMV compared with TMV might constitute an N-terminal region which tends to bury the C-terminal end, but the ROA spectra suggest that this region is not organized into a $\beta$-sheet as proposed by Baratova *et al.* (1992a). In this connection it is worth noting that there is no similarity in primary sequence of PVX and TMV coat proteins since no alignment of the potexvirus and tobamovirus conserved primary sequences is possible and also that, according to a prediction of Sawyer *et al.* (1987), the N-terminal 50 residues of PVX have zero probability of being $\alpha$-helix. This last prediction is unfortunate since it is tempting to assign the greater hydrated $\alpha$-helix content of PVX and NMV relative to TMV to the extra sequences.

TRV exhibits a strong positive ROA band at $\sim$ 1315 cm$^{-1}$ assigned to polyproline II (PPII)-helical structure in the additional central and C-terminal sequences which the coat proteins of this virus contain compared with TMV (Blanch *et al.*, 2001a). However, since there is no additional positive intensity in this region of the ROA spectra of PVX and NMV compared with TMV, the additional sequences of PVX and NMV do not appear to contain much PPII structure. (All three virus ROA spectra in Fig. 1 show similar small but clear positive ROA bands at $\sim$ 1316 cm$^{-1}$ assigned to PPII structure.)

A few ROA bands from side-chains can also be identified. ROA bands in the range $\sim$ 1420–1480 cm$^{-1}$ originate in both aliphatic and aromatic side-chains. Proteins often show a tryptophan ROA band in the $\sim$ 1550 cm$^{-1}$ region, assigned to a W3 type vibration of the indole ring, which reflects the sign and magnitude of the torsion angle $\chi^{2,1}$ (Blanch *et al.*, 2001b). The PVX and NMV strains used here contain five and four tryptophans, respectively. The absence of significant ROA in this spectral region of both viruses therefore suggests conformational heterogeneity among these tryptophans (Blanch *et al.*, 2001b). On the other hand, the two and six tyrosines present in PVX and NMV, respectively, could be responsible for the small but significant negative ROA bands at $\sim$ 1604 cm$^{-1}$ associated with Y8a,b type modes of the aromatic ring (Miura & Thomas, 1995), suggesting these side-chains have mainly fixed conformations generating ROA signals which tend to reinforce. A similar negative band is

present at $\sim$ 1608 cm$^{-1}$ in the ROA spectrum of TMV, which contains four tyrosines. These negative Y8 ROA bands may originate in the participation of some tyrosine side-chains in subunit binding, as observed in the TMV X-ray fibre structure of Namba *et al.* (1989).

## Principal component analysis of coat protein folds

We are developing a pattern recognition program, based on the multivariate statistical approach of principal component analysis (PCA), to identify protein folds from ROA spectral band patterns. The method is similar to one developed for the analysis of conventional Raman spectra of parchment (Nielsen *et al.*, 1999) and is related to methods used for the determination of the structure of proteins from infrared vibrational circular dichroism (Pancoska *et al.*, 1991) and UVCD (Venyaminov & Yang, 1996) spectra. From the ROA spectral data, the PCA algorithm calculates a set of sub-spectra that serve as basis functions, the algebraic combination of which with appropriate expansion coefficients can be used to reconstruct any member of the original set of experimental ROA spectra. The level of accuracy is determined by the number of basis functions used; for example, a set of ten basis functions is sufficient to reproduce adequately any member of our current set of 56 polypeptide, protein and virus ROA spectra. This set contains the ROA spectra of poly(L-lysine) and poly(L-glutamic acid) in model $\alpha$-helical and unordered ('random coil') states, 32 proteins with well-defined tertiary folds known mostly from X-ray crystallography with a few known from multidimensional solution nuclear magnetic resonance, 11 denatured or natively unfolded proteins with structures known to be mostly unordered mainly from spectroscopic techniques such as UVCD, six viruses (including TMV) with coat protein folds known from X-ray crystallography or fibre diffraction and three viruses (PVX, NMV and TRV) with unknown coat protein folds. Since the initial results are promising, we show in Fig. 2 a plot of the coefficients for the whole set of ROA spectra for the two most important basis functions. This provides a two-dimensional representation of the structural relationships among the members of the set. The proteins are colour-coded with respect to the seven different structural types listed on the figure and defined more fully in the caption. These structural types provide a useful initial classification that will be refined and enlarged in later work. Typical examples of these structural types are: alpha, human serum albumin; mainly alpha, hen lysozyme; alpha beta, hen ovalbumin; mainly beta, bovine $\beta$-lactoglobulin; beta, jack bean concanavalin A; mainly unordered, bovine $\beta$-casein; unordered, hen phosvitin. This provides a starting point for the pattern recognition method since it reveals an initial separation of the spectra into clusters corresponding to different dominant types of protein structural elements, with increasing $\alpha$-helix content to the left, increasing $\beta$-sheet content to the right, and increasing disorder from bottom to top. The way in which
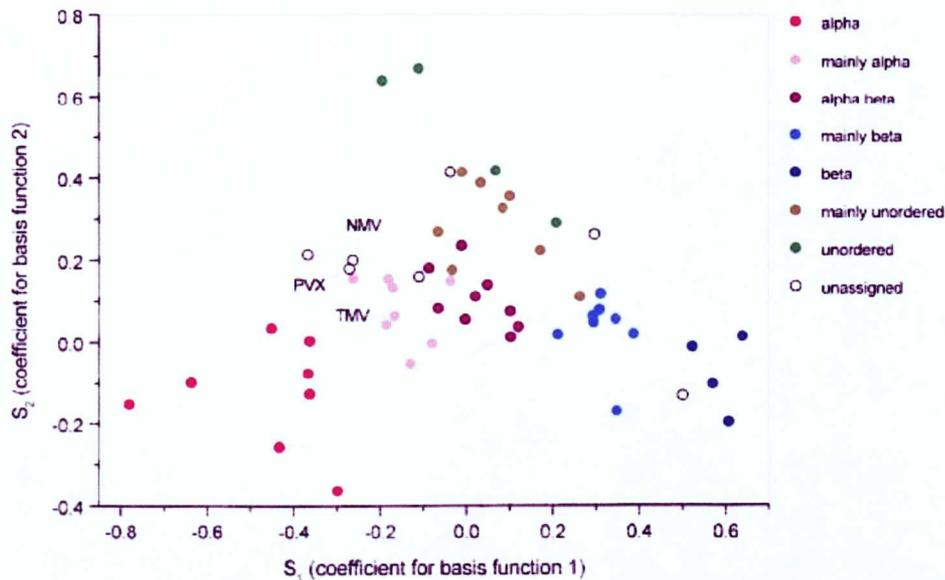
**Fig. 2.** Plot of the PCA coefficients for basis function 1 versus basis function 2 for a set of 56 polypeptide, protein and virus ROA spectra. PVX, NMV and TMV cluster together within the mainly alpha region. More complete definitions of the structural types are as follows: alpha, $> \sim 60\%$ $\alpha$-helix with the rest mainly unordered and little or no $\beta$-sheet; mainly alpha, $> \sim 35\%$ $\alpha$-helix and a small amount of $\beta$-sheet ($\sim 5$–$15\%$); alpha beta, similar significant amounts of $\alpha$-helix and $\beta$-sheet; mainly beta, $> \sim 35\%$ $\beta$-sheet and a small amount of $\alpha$-helix ($\sim 5$–$15\%$); beta, $> \sim 45\%$ $\beta$-sheet with the rest mainly unordered and little or no $\alpha$-helix.

PVX, NMV and TMV cluster together suggests that the folds of their coat proteins may be similar, but with those of PVX and NMV more similar to each other (since their positions are almost identical) than to that of TMV (which is shifted a little from the other two). This cluster lies within the mainly alpha region and so reinforces our conclusion above that the model of Baratova *et al.* (1992a) for the PVX coat proteins contains too much $\beta$-sheet. Plots of other coefficients, which will be given in a later paper, provide further discrimination between different structural types. We have shown these preliminary results here because this unbiased mathematical analysis of the ROA spectral data reinforces the visual impression of a close similarity between the coat protein folds of PVX, NMV and TMV, and because pattern recognition methods are expected to become increasingly important in future applications of ROA to structural virology.

### The overall structures of PVX and NMV

Despite the lack of significant amino acid sequence homology to the TMV coat protein, Sawyer *et al.* (1987) suggested on the basis of predicted $\alpha$-helical segments that the PVX coat protein exhibits an approximate twofold symmetry between the amino and carboxyl halves of the molecule such that the $\alpha$-helices align themselves in pairs like in the TMV coat protein. This reinforces our conclusion from the ROA data that the PVX, NMV and TMV coat protein folds are similar. Furthermore, the low resolution X-ray fibre diffraction data

available for several potexviruses are consistent with $\alpha$-helices packed approximately at right angles to the viral axis, as in TMV (Stubbs, 1989). Hence the helical architecture of the PVX and NMV virus particles may be similar to that of TMV, but with different structural parameters. These different structural parameters may be associated with the small differences in the coat protein folds of PVX and NMV compared with TMV indicated by this study and which, together with their extra sequences, would lead to differences in packing of the protein subunits.

## References

**Abou Haidar, M. G., Xu, H. & Hefferon, K. L. (1998).** Potexvirus isolation and RNA extraction. In *Plant Virus Protocols*, pp. 131–143. Edited by G. D. Foster & S. C. Taylor. Totowa, NJ: Humana Press.

**Baratova, L. A., Grebenshchikov, N. I., Dobrov, E. N., Gedrovich, A. V., Kashirin, I. A., Shishkov, A. V., Efimov, A. V., Järvekülg, L., Radavsky, Y. L. & Saarma, M. (1992a).** The organization of potato virus X coat proteins in virus particles studied by tritium planigraphy and model building. *Virology* **188**, 175–180.

**Baratova, L. A., Grebenshchikov, N. I., Shishkov, A. V., Kashirin, I. A., Radavsky, J. L., Järvekülg, L. & Saarma, M. (1992b).** The topography of the surface of potato virus X: tritium planigraphy and immunological analysis. *Journal of General Virology* **73**, 229–235.

Barron, L. D. & Hecht, L. (2000). Vibrational Raman optical activity: from fundamentals to biochemical applications. In *Circular Dichroism. Principles and Applications*, pp. 667–701. Edited by N. Berova, K. Nakanishi & R. W. Woody. New York: John Wiley & Sons.

Barron, L. D., Hecht, L., Blanch, E. W. & Bell, A. F. (2000). Solution structure and dynamics of biomolecules from Raman optical activity. *Progress in Biophysics & Molecular Biology* **73**, 1–49.

Bell, A. F., Hecht, L. & Barron, L. D. (1997). Vibrational Raman optical activity as a probe of polyribonucleotide stereochemistry. *Journal of the American Chemical Society* **119**, 6006–6013.

Bell, A. F., Hecht, L. & Barron, L. D. (1998). Vibrational Raman optical activity of DNA and RNA. *Journal of the American Chemical Society* **120**, 5820–5821.

Blanch, E. W., Bell, A. F., Hecht, L., Day, L. A. & Barron, L. D. (1999). Raman optical activity of filamentous bacteriophages: hydration of α-helices. *Journal of Molecular Biology* **290**, 1–7.

Blanch, E. W., Robinson, D. J., Hecht, L. & Barron, L. D. (2001a). A comparison of the solution structures of tobacco rattle and tobacco mosaic viruses from Raman optical activity. *Journal of General Virology* **82**, 1499–1502.

Blanch, E. W., Pederson, D. M., Hecht, L., Day, L. A. & Barron, L. D. (2001b). Tryptophan absolute stereochemistry in viral coat proteins from Raman optical activity. *Journal of the American Chemical Society* **123**, 4863–4864.

Brunt, A. A., Foster, G. D., Morozov, S. Y. & Zavriev, S. K. (2000). Genus *Potexvirus*. In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 975–981. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Carey, P. R. (1982). *Biochemical Applications of Raman and Resonance Raman Spectroscopies*. New York: Academic Press.

Goldanskii, V. I., Kashirin, I. A., Shishkov, A. V., Baratova, L. A. & Grebenshchikov, N. I. (1988). The use of thermally activated tritium atoms for structural-biological investigations: the topography of the TMV protein-accessible surface of the virus. *Journal of Molecular Biology* **201**, 567–574.

Goodman, R. M. (1973). Purification of potato virus X and the molecular weight of its coat protein subunits. *John Innes Institute, 64th Annual Report*, 126.

Hecht, L., Barron, L. D., Blanch, E. W., Bell, A. F. & Day, L. A. (1999). Raman optical activity instrument for studies of biopolymer structure and dynamics. *Journal of Raman Spectroscopy* **30**, 815–825.

Homer, R. B. & Goodman, R. M. (1975). Circular dichroism and fluorescence studies on potato virus X and its structural components. *Biochimica et Biophysica Acta* **378**, 296–304.

Jones, R. A. C. (1982). Breakdown of potato virus X resistance gene Nx: selection of a group four strain from strain group three. *Plant Pathology* **31**, 325–331.

Miura, T. & Thomas, G. J., Jr (1995). Raman spectroscopy of proteins and their assemblies. In *Subcellular Biochemistry*, vol. 24, *Proteins: Structure, Function, and Engineering*, pp. 55–99. Edited by B. B. Biswas & S. Roy. New York: Plenum Press.

Nafie, L. A. (1997). Infrared and Raman vibrational optical activity. *Annual Review of Physical Chemistry* **48**, 357–386.

Namba, K., Pattanayek, R. & Stubbs, G. (1989). Visualization of protein–nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2·9 Å by X-ray fiber diffraction. *Journal of Molecular Biology* **208**, 307–325.

Nielsen, K., Garp, T., Paulsen, A. & Boghosian, S. (1999). Principal component analysis of Raman spectra of parchment. In *Proceedings of the International George Papatheodorou Symposium*, pp. 244–248. Edited by S. Boghosian, V. Dracopoulos, C. G. Kontoyannis & G. A. Voyiatzis. Greece: Patras Science Park.

Pancoska, P., Yasui, S. C. & Keiderling, T. A. (1991). Statistical analysis of the vibrational circular dichroism of selected proteins and relationship to secondary structure. *Biochemistry* **30**, 5089–5103.

Sawyer, L., Tollin, P. & Wilson, H. R. (1987). A comparison between the predicted secondary structures of potato virus X and papaya mosaic virus coat proteins. *Journal of General Virology* **68**, 1229–1232.

Stubbs, G. (1989). Virus structure. In *Prediction of Protein Structure and Principles of Protein Conformation*, pp. 117–148. Edited by G. Fasman. New York: Plenum Press.

Stubbs, G. (1999). Tobacco mosaic virus particle structure and the initiation of disassembly. *Philosophical Transactions of the Royal Society B* **354**, 551–557.

Thomas, G. J., Jr (1987). Viruses and nucleoproteins. In *Biological Applications of Raman Spectroscopy*, vol. 1, pp. 135–201. Edited by T. G. Spiro. New York: John Wiley & Sons.

Thomas, G. J., Jr (1999). Raman spectroscopy of protein and nucleic acid assemblies. *Annual Review of Biophysics and Biomolecular Structure* **28**, 1–27.

Tollin, P., Wilson, H. R. & Bancroft, J. B. (1980). Further observations on the structure of particles of potato virus X. *Journal of General Virology* **49**, 407–410.

Venyaminov, S. Y. & Yang, J. T. (1996). Determination of protein secondary structure. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, pp. 69–107. Edited by G. D. Fasman. New York: Plenum Press.

Wilson, H. R., Tollin, P., Sawyer, L., Robinson, D. J., Price, N. C. & Kelly, S. M. (1991). Secondary structures of narcissus mosaic virus coat protein. *Journal of General Virology* **72**, 1479–1480.

# References

Acharya, K. R.; Stuart, D. I.; Walker, N. P.; Lewis, M.; Phillips, D. C., *J. Mol. Biol.*, **208**, 99 (1989)

Adzhubei, A. A.; Eisenmenger, F.; Tumanyan, V. G.; Zinke, M.; Brodzinski, S.; Esipova, N. G. *J. Biomol. Struct. Dynam.*, **5**, 689-704 (1987)

Adzhubei, A. A.; Sternberg, M. J. E., *J. Mol. Biol.*, **229**, 472-493 (1993)

Athès, V.; Combes, D.; Zwick, A. *J. Raman Spectrosc.*, **29**, 373-378 (1998)

Atkins, P. W.; Barron, L. D., *Mol. Phys.*, **16**, 453-466 (1969)

Banaszak. L.; Winter, N.; Xu, Z.; Bernlohr, D. A.; Cowan, S.; Jones, T. A., *Adv. Protein. Chem.*, **45**, 89-151 (1994)

Baratova, L. A.; Grebenshchikov, N. I.; Dobrov, E. N.; Gedrovich, A. V.; Kashirin, I. A.; Shishkov, A. V.; Efimov, A. V.; Järvekülg, L.; Radavsky, Y. L.; Sarama, M., *Virology*, **188**, 175-180 (1992)

Barghorn, S.; Zheng-Fischhöfer, Q.; Ackmann, M.; Biernat, J.; von Bergen, M.; Mandelkow, E. M.; Mandelkow, E., *Biochemistry*, **39**, 11714-11721 (2000)

Barron, L. D., *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge (1982)

Barron, L. D.; Buckingham, A. D.; *Mol. Phys.*, **20**, 1111-1119 (1971)

Barron, L. D.; Buckingham, A. D, *J. Am. Chem. Soc.*, **96**, 4769-4773 (1974)

Barron, L. D.; Boogard, M. P.; Buckingham, A. D., *J. Am. Chem. Soc.*, **95**, 603-605 (1973)

Barron, L. D.; Hecht, L., In *Circular Dichroism: Principles and Applications*, 667-701. Edited by N. Berova, K. Nakanishi & R. W. Woody, John Wiley & Sons, New York (2000)

Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F., *Prog. Biophys. Mol, Biol.*, **73**, 1-49 (2000)

Benyon, J. D. E.; Lamb, D. R., *Charge-Coupled Devices and Their Applications*, McGraw-Hill, London (1980)

Biere, A. L.; Wood, S. J.; Wypych, J.; Steavenson, S.; Jiang, Y.; Anafi, D.; Jacobsen, F. W.; Jarosinski, M. A.; Wu, G. M.; Louis, J. C.; Martin, F.; Narhi, L. O.; Citron, M., *J. Biol. Chem.*, **275**, 34574-34579 (2000)

Birktoft, J. J.; Blow, D. M., *J. Mol. Biol.*, **68**, 187 (1972)

Blanch, E. W.; Bell, A. F.; Hecht, L.; Day, L. A.; Barron, L. D., *J. Mol. Biol.*, **290**, 1-7, (1999)

Blanch, E. W.; Hecht, L., Barron, L. D., *Protein Science*, **8**, 1362-1367 (1999)

Blanch, E. W.; Morozova-Roche, L. A.; Cochran, D. A. E. ; Doig, A. J.; Hecht, L.; Barron, L. D., *J. Mol. Biol.*, **301**, 553-563 (2000a)

Blanch, E. W.; Morozova-Roche, L. A.; Hecht, L., Noppe, W.; Barron, L. D., *Biopolymers (Biospectroscopy)*, **57**, 235 -248 (2000b)

Blanch, E. W.; Pederson, D. M.; Hecht, L.; Day, L. A.; Barron, L. D., *J. Am. Chem. Soc.*, **123**, 4863-4864 (2001)

Blanch, E. W.; Robinson, D. J.; Hecht, L.; Barron, L. D., *J. Gen. Virology*, **82**, 1499-1502 (2001)

Blanch, E. W.; Robinson, D. J.; Hecht, L.; Syme, C. D.; Nielsen, K.; Barron, L. D., *J. Gen. Virology*, **83**, in press (2002)

Bode, W.; Epp, O.; Huber, R.; Laskowski Jr., M.; Ardelt, W., *Eur. J. Biochem.*, **147**, 387 (1985)

Branden, C.; Tooze, J., *Introduction to Protein Structure*, Garland Publishing, New York (1991)

Braun, W.; Vašák, M.; Robbins, A. H.; Stout, C. D.; Wagner, G.; Kägi, J. H. R.; Wüthrich, K., *Proc. Natl. Acad. Sci. USA*, **89**, 10124-10128 (1992)

Brownlow, S.; Morais-Cabral, J. H.; Cooper, R.; Flower, D. R.; Yewdall, S. J.; Polikarpov, I.; North, A. C.; Sawyer, L., *Structure*, **5**, 481 (1997)

Buckingham, A. D.; *Adv. Chem. Phys.*, **12**, 107-142 (1967)

Burley, R. W; Vadhera, D. V., *The Avain Egg:Chemsitry and Biology*; Wiley: New York, (1989)

Byler, D. M.; Susi, H., *J. Indust. Microbiol.*, **3**, 73-88 (1988)

Carter, D. C.; Ho, J. X., *Adv. Protein Chem.*, **45**, 153-203 (1994)

Castano, E. M.; Prelli, F. C.; Frangione, B., *Lab. Invest.*, **73**, 457-460 (1995)

Chen, P.; Rose, J.; Love, R.; Wei, C. H.; Wang, B. C., *J. Biol. Chem.*, **267**, 1990-1994 (1992)

Creamer, L. K.; Richardson, T.; Parry, D. A. D., *Arch. Biochem. Biophys.*, **211**, 689-696 (1981)

Curry, S.; Mandelkow, H.; Brick, P.; Franks, N., *Nat. Struct. Biol.*, **5**, 827 (1998)

Dawson, D.; Eppenberger, H.; Kaplan, N., *J. Biol. Chem.*, **242**, 210 (1967)

de Gruyter, W., *Concise Encyclopedia Biochemistry, Second Edition*, New York
(1988)

de la Sierra, I. L.; Quillien, L.; Flecker, P.; Gueguen, J.; Brunie, S., *J. Mol. Biol.*, 285,
1195-1207 (1999)

Diem, M., *Modern Vibrational Spectroscopy*, Wiley, New York (1993)

Dukor R. K.; Keiderling, T. A., *Biopolymers*, 31, 1747- 1761 (1991)

Edmundson, A. B.; Ely, K. R.; Sly, D. A.; Westholm, F. A.; Powers, D. A.; Liener, L.
E. *Biochemistry*, 10, 3554 (1971)

Gamblin, S. J.; Cooper, B.; Millar, J. R.; Davies, G. J.; Littlechild, J. A.; Watson, H.
C., *FEBS Letters*, 262, 282 (1990)

Gargaro, A. R.; Barron, L. D.; Hecht, L., *J. Raman Spectrosc.*, 24, 91-96, (1993)

Goedert, M., *Phil. Trans. R. Soc. Lond.* B, 354, 1101-1118 (1999)

Goedert, M.; Spillantini, M. G.; Jakes, R.; Rutherford, D.; Crowther, R. A., *Neuron*, 3,
519-526 (1989)

Graham, E. R. B.; Malcolm, G. N.; McKenzie, H. A., *Int. J. Biol. Macromol.,* 6, 155-
161 (1984)

Griffin, M. C. A.; Price, J. C.; Martin, S. R., *Int. J. Biol. Macromol.*, 8, 367-371
(1986)

Groves, M. L.; Dower, H. J.; Farrell, H. M. Jr, *J. Prot. Chem.*, 11, 21-28 (1992)

Harata, K.; Abe, Y.; Muraki, M., *Proteins*, 30, 232 (1998)

Hecht, L.; Barron, L. D., *Appl. Spectrosc.*, 44, 483-491 (1990)

Hecht, L.; Barron, L. D., *Faraday Discuss.*, 99, 35-47 (1994)

Hecht, L.; Barron, L. D.; *J. Mol. Structure*, 347, 449-458 (1995)

Hecht, L.; Barron, L. D.; Blanch, E. W.; Bell, A. F.; Day, L. A., *J. Raman Spectrosc.*,
30, 815-825 (1999)

Hecht, L.; Barron, L. D.; Gargaro, A. R.; Wen, Z. Q.; Hug, W., *J. Raman Spectrosc.*,
23, 401-411 (1992)

Hecht, L.; Barron, L. D.; Hug., W., *Chem. Phys. Lett.*, 158, 341-344 (1989)

Hohenester, E.; Hutchinson, W. L.; Pepys, M. B.; Wood, S. P., *J. Mol. Biol.*, 269,
3778-3786 (1997)

Holt, C., *Adv. Protein Chem.*, 43, 63-151 (1992)

Holt, C.; Sawyer, L., *J. Chem. Soc. Faraday Trans.*, 89, 2683-2692 (1993)

Hoogstraten, C.G.; Choe, S.; Westler, W. M.; Markley, J. L., *Protein* Science, 4,
2289-2299 (1995)

Jakes, R.; Spillantini, M. G.; Goedert, M., *FEBS Letters*, **345**, 27-32 (1994)

Ji, H.; Liu, Y. E.; Jia, T.; Wang, M.; Liu, J.; Xiao, G.; Joseph, B. K., Rosen, C.; Shi, Y. E., *Cancer Res.*, **57**, 759-764 (1998)

Kabsch, W.; Sander, C., *Biopolymers*, **22**, 2577-2637 (1983)

Kossiakoff, A. A.; Chambers, J. L.; Kay, L. M.; Stroud, R. M., *Biochemistry*, **16**, 654 (1977)

Kraulis, P. J., *J. Appl Cryst.*, **24**, 946-950 (1991)

Kremer, J. M. H.; Wilting, J.; Janssen, L. H. M., *Pharmacol. Rev.*, 1-47 (1988)

Krimm, S.; Bandekar, J., *Biopolymers*, **19**, 1-29 (1980)

Krimm, S.; Bandekar, J., *Adv. Protein Chem.*, **38**, 181-364 (1986)

Kurinov, I. V.; Harrison, R. W., *Acta Crystallogr. D Biol. Crystallogr.*, **51**, 98 (1995)

Linderstrøm-Lang, K. U., *The Lane Medical Lectures*, Stanford University Press, Stanford, California (1952)

Losso, J. N; Bogumil, R.; Nakai, S., *Comp. Biochem. Physiol.*, **106B**, 919-923 (1993)

Messerle, B. A.; Schäffer, A.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K., *J. Mol. Biol.*, **214**, 765-779 (1990)

Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C., *J. Mol. Biol.*, **247**, 536-540 (1995)

Nafie, L.A.; Che, D., *Adv. Chem. Phys.*, **85**, 105-222 (1994)

Neidhart, D. J.; Petsko, G. A., *Protein Eng.*, **2**, 271 (1988)

Nesloney, C. L.; Kelly, J. W., Bioorg. Med. Chem., **4**, 739-766 (1996)

Nielsen, K.; Garp, T.; Paulsen, A.; Boghosian, S., In *Proceedings of the International George Papatheodorou Symposium*, 244-248. Edited by S. Boghosian, V. Dracopoulos, C. G. Kontoyannis & G. A. Voyiatis, Patras Science Park, Greece (1999)

Ono, T.; Yada, R.; Yutani, K.; Nakai, S., *Biochim. Biophys. Acta*, **911**, 318-325 (1987)

Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M., *Structure*, **5**, 1093-1108 (1997)

Orengo, C. A.; Jones, D. T.; Thornton, J. M., *Nature*, **372**, 631-634 (1994)

Pancoska, P., In *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, New York (2000)

Pancoska, P.; Yasui, S. C.; Keiderling, T. A., *Biochemistry*, **30**, Issue 20, 5089-5103 (1991)

Pande, J.; Pande, C.; Gilg, D.; Vašák, M.; Callender, R.; Kägi, J. H. R., *Biochemistry*, **25**, 5526-5532 (1986)

Pauling, L.; Corey, R. B., *Proc. Natl. Acad. Sci. U.S.A.*, **37**, 205-211 (1951)

Pauling, L.; Corey, R. B.; Branson, H. R., *Proc. Natl. Acad. Sci. U.S.A.*, **37**, 729-740 (1951)

Pepys, M. B.; Booth, D. R.; Hutchison, W. L.; Gallimore, J. R.; Collins, P. M.; Hohenester, E., *Amyloid: Int. J. Exp. Clin. Invest.*, **4**, 274-295 (1997)

Perutz, M., *Protein Structure*, W. H. Freeman, New York (1992)

Pervaiz, S.; Brew, K., *FASEB J.*, **1**(3), 209-14 (1987)

Poland, D.; Scheraga, H. A., *Theory of Helix-Coil Transitions in Biological Macromolecules*, Academic Press, New York (1970)

Polavarapu, P. L., *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*, Elsevier, Amsterdam (1998)

Polavarapu, P. L.; Deng, Z., *Faraday Discuss.*, **99**, 151-163 (1994)

Pons, T.: Olmea, O; Chinea, G.; Beldarraín, A.; Márquez, G.; Acosta, N.; Rodríguez, L.; Valencia, A., *Proteins Struct. Funct. Genet.*, **33**, 383-395, (1998)

Prescott, B.; Renugopalakrishnan, V.; Glimcher, M. J.; Bhushan, A; Thomas, G. J., *Biochemistry*, **25**, 2792-2798 (1986)

Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M., *J. Mol. Biol.*, **178**, 787 (1984)

Qin, B. Y.; Bewley, M. C.; Creamer, L. K.; Baker, H. M.; Baker, E. N.; Jameson, G. B., *Biochemistry*, **37**, 14014-14023 (1998)

Raap, J.; Kerling, K. E. T.; Vreeman, H.; Visser, S., *Arch. Biochem. Biophys.*, **221**, 117-124 (1983)

Ramakrishnan, C.; Ramachandran, G. N., *Biophysical Journal*, **5**, Issue 6, 909-933 (1965)

Rao, J. K.; Bujacz, G.; Wlodawer, A., *FEBS Letters*, **439**, 133 (1998)

Rasmussen, L. K.; Højrup, P.; Petersen, T. E., *Eur. J. Biochem.*, **203**, 381-386 (1992a)

Rasmussen, L. K.; Højrup, P.; Petersen, T. E., *Eur. J. Biochem.*, **207**, 215-222 (1992b)

Rawlings, N.; Barrett, A. J., *Methods Enzymol.*, **244**, 19- 61 (1994)

Reeke, G. N.; Becker, J. W.; Edelman, G. M., *J. Biol. Chem.*, **250**, 1525-1547 (1975)

Robbins, A. H.; McRee, D. E.; Williamson, M.; Collett, S. A., Xuong, N. H.; Furey, W. F.; Wang, B. C.; Stout, C. D., *J. Mol. Biol.*, **221**, 1269-1293 (1991)

Rogers, A.; Norden, B., *Circular Dichroism and Linear Dichroism*, Oxford Chemistry Masters, Oxford University Press (1997)

Rollema, H. S., In *Advanced Dairy Chemistry-I: Proteins* (Fox, P. F., ed) 111-140. Elsevier Applied Science, London (1992)

Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek R. A., *Biochemistry*, **33**, 17-22 (1994)

Rypniewski, W. R.; Dambmann, C.; Vonderosten, C.; Dauter, M.; Wilson, K. S.; *Acta Crystallogr. D Biol. Crystallogr.*, **51**, 73 (1995)

Salemme, F. R., Prog. Biophys. Mol. Biol., **42**, 95-133 (1983)

Saul, F.A; Poljack, R. J., *Proteins*, **14**, 363 (1991)

Sawyer, L., *Nature*, **327**, 659 (1987)

Sawyer, L.; Tollin, P.; Wilson, H. R., *J. Gen. Virology*, **68**, 1229-1232 (1987)

Schülke, N.; Schmid, F. X., *J. Biol. Chem.*, **263**, 8832-8837 (1988)

Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K., *J. Mol. Biol.*, **203**, 251-268 (1988)

Schweers, O.; Schönbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E., *J. Biol. Chem.*, **269**, 24290-24297 (1994)

Serpell, L. C.; Berriman, J.; Jakes, R.; Goedert, M.; Crowther, R. A., *Proc. Natl. Acad. Sci. USA*, **87**, 4897-4902

Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M., *Fold. Des*, **1**, R95-R106 (1996)

Smyth, E.; Syme, C. D.; Blanch, E. W.; Hecht, L.; Vašák, M.; Barron, L. D., *Biopolymers*, **58**, 138-151 (2001)

Sönnichsen, F. D.; Chao, H. L.; Davies, P. L.; Sykes, B. D., *Science*, **259**, 1154-1157 (1993)

Spillantini, M. G.; Schmidt, M. L.; Lee, V. M. Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M., *Nature*, **388**, 839-840 (1997)

Stein, P. E.; Leslie, A. G.; Finch, J. T.; Carrell, R. W.; *J. Mol. Biol.*, **221**, 941 (1991)

Sweedler, J. V.; Bilhorn, R. B.; Epperson, P. M.; Sims, G. R.; Denton, M. B., *Anal. Chem.*, **60**, 282A (1988)

Swift, H. J.; Brady, L.; Derewenda, Z. S.; Dodson, E. J.; Dodson, G. G.; Turkenburg, J. P.; Wilkinson, A. J., *Acta Crystallogr. B*, **47**, 535 (1991)

Syme, C. D.; Blanch, E. W.; Holt, C.; Jakes, R.; Goedert, M.; Hecht, L.; Barron, L. D., *Eur. J. Biochem.*, **269**, 148-156 (2002)

Tanford, C.; Bunville, L. G.; Nozaki, Y., *J. Am. Chem. Soc.*, **81**, 4032-4036 (1959)

Tedesco, J. M.; Owen, H.; Pallister, D. M.; Morris, M. D., *Anal. Chem.*, **65**, 441A-449A (1993)

Thornton, J. M.; Jones, D. T.; MacArthur, M. W.; Orengo C. A.; Swindells, M. B., *Phil. Trans. R. Soc. Lond. B*, **348**, 71-79 (1995)

Tiffany, M. L.; Krimm, S., *Biopolymers*, **11**, 2309-2316 (1972)

Torreggiani, A.; Fini, G., *J. Raman Spectrosc.*, **29**, 229-236 (1998)

Tsuge, H.; Ago, H.; Noma, M.; Nitta, K.; Sugai, S.; Miyano, M., *J. Biochem. (Tokyo)*, **111**, 141 (1992)

Uéda, K.; Fukushima, H.; Masliah, E.; Xia, Y.; Iwai, A.; Yoshimoto, M.; Otero, D. A. C.; Kondo, J.; Ihara, Y.; Saitoh, T., *Proc. Natl. Acad. Sci. USA*, **90**, 11282-11286 (1993)

Venyaminov, S. Y.; Yang, J. T., In *Circular Dichroism and the Conformational Analysis of Biomolecules*, 69-107. Edited by G. D. Fasman, Plenum Press, New York (1996)

Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J., *J. Mol. Biol.*, **194**, 531-544 (1987)

Wang, J. L.; Cunningham, B. A.; Edelman, G. M. *Proc. Natl. Acad. Sci USA*, **68** 1130-1134 (1971)

Vogel, H. J., *Biochemistry*, **22**, 668-674 (1983)

Watanabe, K.; Matsuda, T.; Sato, Y., *Biochim. Biophys. Acta*, **667**, 242 (1981).

Watts, D., in *The Enzymes Vol. VIII*, Pt. A, 3rd Ed., (Boyer, P.D., ed.), Academic Press, New York, 383 (1973)

Weber, E.; Papamokos, E.; Bode, W.; Huber, R.; Kato, I.; Laskowski Jr., M., *J. Mol. Biol.*, **149**, 109-123 (1981)

Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T. Jr., *Biochemistry*, **35**, 13709-13715 (1996)

Weisgerber, S.; Helliwell, J. R., *J. Chem. Soc. Faraday Trans.*, **89**, 2667-2675 (1993)

Werner, M. H.; Wemmer, D. E., *Biochemistry*, **31**, 999-1010 (1992)

Whittingham, J. L.; Edwards, D. J.; Antson, A. A.; Clarkson, J. M.; Dodson, G. G., *Biochemistry*, **37**, 11516 (1998)

Wiborg, O.; Pedersen, M. S.; Wind, A.; Berglund, L. E.; Marcker, K. A.; Vuust, J., *EMBO J.*, **4**, 755-759 (1985)

Wilson, E. B.; Decius, J. C.; Cross, P. C., *Molecular Vibrations*, McGraw-Hill, New York (1955)

Wilson, G., *Doctoral thesis, University of Glasgow* (1996)

Woody, R. W., *Adv. Biophys. Chem.*, **2**, 37-79 (1992)

Woody, R. W., In *Circular Dichroism: Principles and Applications*, 601-620, Edited by N. Berova, K. Nakanishi & R. W. Woody, John Wiley & Sons, New York (2000)

Yasui, S. C.; Pancoska, P.; Dukor, R. K.; Keiderling, T. A.; Renugopalakrishnan, V.; Glimcher, M. J.; Clark, R. C., *J. Biol. Chem.*, **265**, 3780-3788 (1990)