

Structure and behaviour of proteins, nucleic acids and viruses from vibrational Raman optical activity

Laurence D. Barron^{a,*}, Ewan W. Blanch^a, Iain H. McColl^a, Christopher D. Syme^a, Lutz Hecht^a and Kurt Nielsen^b

^a *Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK*

^b *Department of Chemistry, DTU 207, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Abstract. On account of its sensitivity to chirality Raman optical activity (ROA), which may be measured as a small difference in vibrational Raman scattering from chiral molecules in right- and left-circularly polarized incident light, is a powerful probe of structure and behaviour of biomolecules in aqueous solution. Protein ROA spectra provide information on the secondary and tertiary structure of the polypeptide backbone, hydration, side chain conformation and structural elements present in denatured states. Nucleic acid ROA spectra provide information on the sugar ring conformation, the base stacking arrangement and the mutual orientation of the sugar and base rings around the C–N glycosidic link. The ROA spectra of intact viruses provide information on the folds of the coat proteins and the nucleic acid structure. The large number of structure-sensitive bands in protein ROA spectra is especially favourable for fold determination using pattern recognition techniques. This article gives a brief account of the ROA technique and presents the ROA spectra of a selection of proteins, nucleic acids and viruses that illustrate the applications of ROA spectroscopy in biomolecular research.

Keywords: Raman optical activity, vibrational spectroscopy, chirality, protein conformation, fold recognition, conformational disease, nucleic acid structure, virus structure

1. Introduction

The techniques of X-ray crystallography and high resolution nuclear magnetic resonance (NMR) spectroscopy currently dominate the field of structural biology due to their ability to reveal the details of biomolecular structure at atomic resolution and will continue to be invaluable. However, there are limitations to their applicability. For example, many biomolecules are difficult to crystallize while many others have structures too large to be solved by current NMR methods; yet academic and commercial interests in the post-genomic era will demand structural information about such biomolecules. While not providing information at atomic resolution, vibrational spectroscopic techniques will become increasingly valuable since they can be routinely applied to a wide range of biological systems and provide information on structure, dynamics, local environments and assembly processes [1–4].

A promising vibrational spectroscopic technique for the study of biomolecules is Raman optical activity (ROA), a novel form of chiroptical spectroscopy which measures a small difference in the intensity of vibrational Raman scattering from chiral molecules in right- and left-circularly polarized incident light or, equivalently, a small circularly polarized component in the scattered light [3–6]. This was first observed by Barron et al. in 1973 [7] and has been developed in our laboratory to the point that it is now an

*Corresponding author. Tel.: +44 141 330 5168; Fax: +44 141 330 4888; E-mail: laurence@chem.gla.ac.uk.

incisive probe of the structure and behaviour of most biomolecules in aqueous solution [8]. The power of ROA in this area derives from the fact that, like the complementary technique of vibrational circular dichroism (VCD) [3,4,6,9], it is a form of vibrational optical activity and so is sensitive to chirality associated with all the $3N-6$ fundamental molecular vibrational transitions, where N is the number of atoms. In this article we provide a brief survey of the ROA technique, with some typical applications in biomolecular science illustrated by examples from recent studies on proteins, nucleic acids and viruses.

2. Basic theory

2.1. The ROA observables

The fundamental scattering mechanism responsible for ROA was discovered by Atkins and Barron [10]. The definitive theory was developed by Barron and Buckingham [11], who introduced the following definition of the dimensionless circular intensity difference (CID),

$$\Delta = (I^R - I^L)/(I^R + I^L)$$

as an appropriate experimental observable, where I^R and I^L are the scattered intensities in right- and left-circularly polarized incident light, respectively. 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 $A_{\alpha\beta\gamma}$ [12,13], the CIDs for forward (0°) and backward (180°) scattering from an isotropic sample for incident wavelengths much greater than the molecular dimensions are as follows:

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

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

where the isotropic invariants are defined as

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

and the anisotropic invariants as

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

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

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

These results apply specifically to Rayleigh (elastic) scattering. For Raman (inelastic) scattering the same basic CID expressions apply but with the molecular property tensors replaced by corresponding vibrational Raman transition tensors. For the case of a molecule composed entirely of idealized axially-symmetric bonds, for which $\beta(G')^2 = \beta(A)^2$ and $\alpha G' = 0$ [13,14], a simple bond polarizability theory shows that ROA is generated exclusively by anisotropic scattering and the CID expressions reduce to

$$\Delta(0^\circ) = 0,$$

$$\Delta(180^\circ) = \frac{32\beta(G')^2}{c[45\alpha^2 + 7\beta(\alpha)^2]}.$$

Unlike conventional Raman scattering intensities, which are the same in forward and backward directions, ROA intensity is therefore maximized in backscattering and zero in forward scattering. These considerations lead to the important conclusion that backscattering boosts the ROA signal relative to the background Raman intensity and is therefore the best experimental strategy for most ROA studies of biomolecules in aqueous solution [15,16].

2.2. Enhanced sensitivity of ROA to structure and dynamics of chiral biomolecules

The normal vibrational modes of biopolymers can be highly complex as they contain contributions from local vibrational coordinates in both the backbone and the side chains. ROA is able to cut through the complexity of the corresponding vibrational spectra as the largest ROA signals originate from vibrational coordinates which sample the most rigid and chiral parts of the structure. These usually reside within the backbone and give rise to ROA band patterns characteristic of the backbone conformation. In proteins these include bands characteristic of secondary, loop and turn structures. In comparison, the parent conventional Raman spectrum of a protein is dominated by bands arising from the amino acid side chains which may obscure the peptide backbone bands.

The time scale of the Raman scattering event ($\sim 3.3 \times 10^{-14}$ s for a vibration with Stokes wavenumber shift 1000 cm^{-1} excited in the visible) is much shorter than that of the fastest conformational fluctuations. The ROA spectrum is therefore a superposition of 'snapshot' spectra from all the distinct conformations present in the sample at equilibrium. ROA intensity is dependent on absolute chirality and therefore yields a cancellation of contributions from enantiomeric structures which can arise as a mobile structure explores the range of accessible conformations. These factors result in ROA exhibiting an enhanced sensitivity to the dynamic behaviour of biomolecular structure. In contrast, observables that are 'blind' to chirality, such as conventional Raman band intensities, are generally additive and therefore less sensitive to conformational mobility. Ultraviolet circular dichroism (UVCD) also shows an enhanced sensitivity to the dynamics of chiral structures. However, the sensitivity is less than that for ROA due to its dependence on electronic transitions.

3. Instrumental

Since ROA is maximized in backscattering, a backscattering geometry has proven to be essential for the routine measurement of ROA spectra of biomolecules in aqueous solution. The optical layout of our current backscattering ROA instrument is shown in Fig. 1 [17]. A visible argon ion laser beam is weakly focused into the sample solution which is contained in a small rectangular fused quartz cell. The cone of

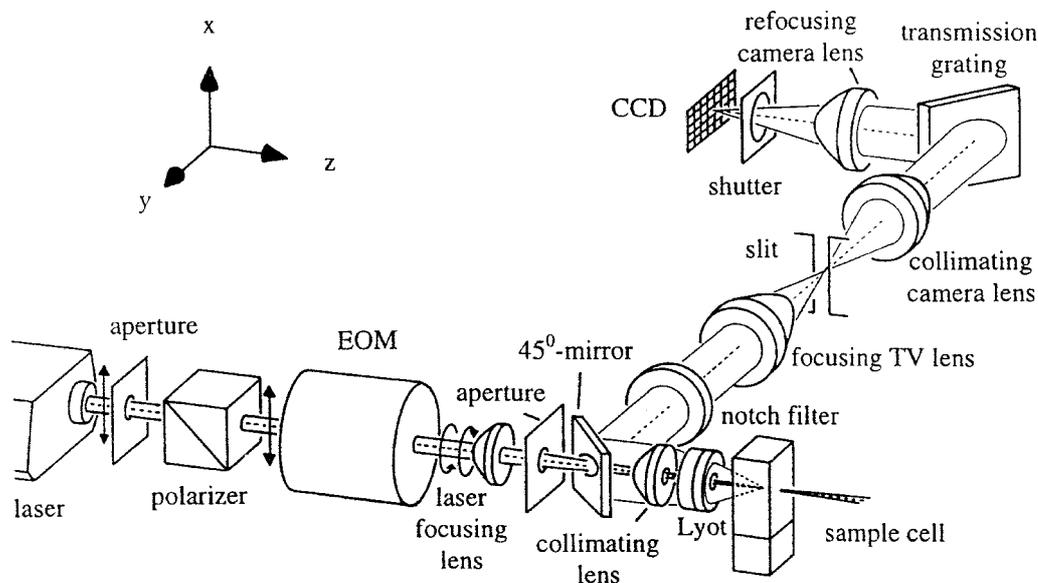


Fig. 1. Layout of the current backscattering ROA instrument used in Glasgow.

backscattered light is reflected off a 45° mirror, which has a small central hole drilled to allow passage of the incident laser beam, through an edge filter to remove the Rayleigh line and into the collection optics of a single grating spectrograph. This is a customized version of a fast imaging spectrograph containing a highly efficient volume holographic transmission grating. The detector is a cooled back-thinned charge coupled device (CCD) camera with a quantum efficiency of $\sim 80\%$ over the spectral range of our studies. Operation of the CCD camera in multichannel mode allows the full spectral range to be measured in a single acquisition. To measure the small ROA signals, the spectral acquisition is synchronized with an electro-optic modulator (EOM) used to switch the state of polarization of the incident laser beam between right and left circular at a suitable rate.

Spectra are displayed in analog-to-digital converter (ADC) units as a function of the Stokes Raman wavenumber shift with respect to the exciting laser line. The green 514.5 nm line of an argon-ion laser is used as this provides a compromise between reduced fluorescence from sample impurities with increasing wavelength and the increased scattering efficiency with decreasing wavelength due to the Rayleigh λ^{-4} law. Typical laser power at the sample is ~ 700 mW and sample concentrations of proteins, polypeptides and nucleic acids are ~ 30 – 100 mg/ml while those of intact viruses are ~ 5 – 30 mg/ml. Under these conditions ROA spectra such as those presented here may be obtained in ~ 5 – 24 hours for proteins and nucleic acids and ~ 1 – 4 days for intact viruses. Measurements can also be performed over the temperature range of ~ 0 – 60°C by directing dry air downwards over the sample cell from a device used to cool protein crystals in X-ray diffraction experiments, in order to study dynamic behaviour.

Several novel design features of a new ROA instrument developed by Hug and Hangartner [18] will be especially valuable for biomolecular studies. It is based on measuring the circularly polarized component in the Raman scattered light instead of the Raman intensity difference in right- and left-circularly polarized incident light employed in the Glasgow instrument. This enables 'flicker noise' arising from dust particles, density fluctuations, etc. to be eliminated since the intensity differences required to extract the circularly polarized components of the scattered beam are taken between two components of the scattered light measured during the same acquisition period. The flicker noise therefore cancels out,

resulting in superior signal-to-noise characteristics. A commercial instrument based on this design will be available shortly from BioTools.

4. Polypeptides and proteins

4.1. General

ROA is an excellent technique for studying polypeptide and protein structure in aqueous solution since, as mentioned above, their ROA spectra are often dominated by bands originating in the peptide backbone which directly reflect the solution conformation. Furthermore, the special sensitivity of ROA to dynamic aspects of structure makes it a new source of information on order–disorder transitions.

Vibrations of the backbone in polypeptides and proteins are usually associated with three main regions of the Raman spectrum [19,20]. These are the backbone skeletal stretch region $\sim 870\text{--}1150\text{ cm}^{-1}$ originating in mainly $C_{\alpha}\text{--C}$, $C_{\alpha}\text{--C}_{\beta}$ and $C_{\alpha}\text{--N}$ stretch coordinates; the amide III region $\sim 1230\text{--}1310\text{ cm}^{-1}$ which is often thought to involve mainly the in-phase combination of largely the N–H in-plane deformation with the $C_{\alpha}\text{--N}$ stretch; and the amide I region $\sim 1630\text{--}1700\text{ cm}^{-1}$ which arises mostly from the C=O stretch. However Diem [3] has shown that, in small peptides, the amide III region involves much more mixing between the N–H and $C_{\alpha}\text{--H}$ deformations than previously thought, and should be extended to at least 1340 cm^{-1} . This extended amide III region is particularly important for ROA studies because, as first shown in an early ROA study of alanyl peptide oligomers [21], the coupling between N–H and $C_{\alpha}\text{--H}$ deformations is very sensitive to geometry and generates a rich and informative ROA band structure. Schweitzer-Stenner [22] has provided a detailed review of the extended amide III and other vibrational modes of small peptides based on conventional Raman studies. Side-chain vibrations also generate many characteristic Raman bands [19,20,23]: although less prominent in ROA spectra, some side-chain vibrations do give rise to useful ROA signals.

Poly(L-lysine) adopts well-defined conformations under certain conditions of temperature and pH and has long been used as a model for the spectroscopic identification of secondary structure sequences in proteins. Poly(L-lysine) at alkaline pH has neutral side chains and so is able to support α -helical conformations stabilized both by internal hydrogen bonds and by hydrogen bonds to the solvent; whereas poly(L-lysine) at neutral and acid pH has charged side chains which repel each other thereby encouraging a disordered structure. The backscattered Raman and ROA spectra of these samples are shown in Figs 2(a) and 2(b), respectively [17,24]. Until very recently it has not been possible to obtain ROA spectra of model β -sheet conformations of any polypeptide due to experimental difficulties associated with the gel-like consistency of such samples at the high concentrations required for ROA measurements. However, thanks to improved instrumentation, this has now been achieved for β -sheet poly(L-lysine), produced by heating the sample at high pH to 50°C for one hour [25]. This β -sheet poly(L-lysine) ROA spectrum is shown for the first time in Fig. 2(c). There are clearly many differences between the ROA spectra of the α -helical, disordered and β -sheet conformations of poly(L-lysine) which enable ROA to easily distinguish between them.

Preliminary accounts of the assignment of polypeptide and protein ROA bands to the various types of structural elements were given in earlier articles [24,26]. However, with the measurement of many more ROA spectra, some of these early assignments have been revised [8] and are still being refined as more data accumulate.

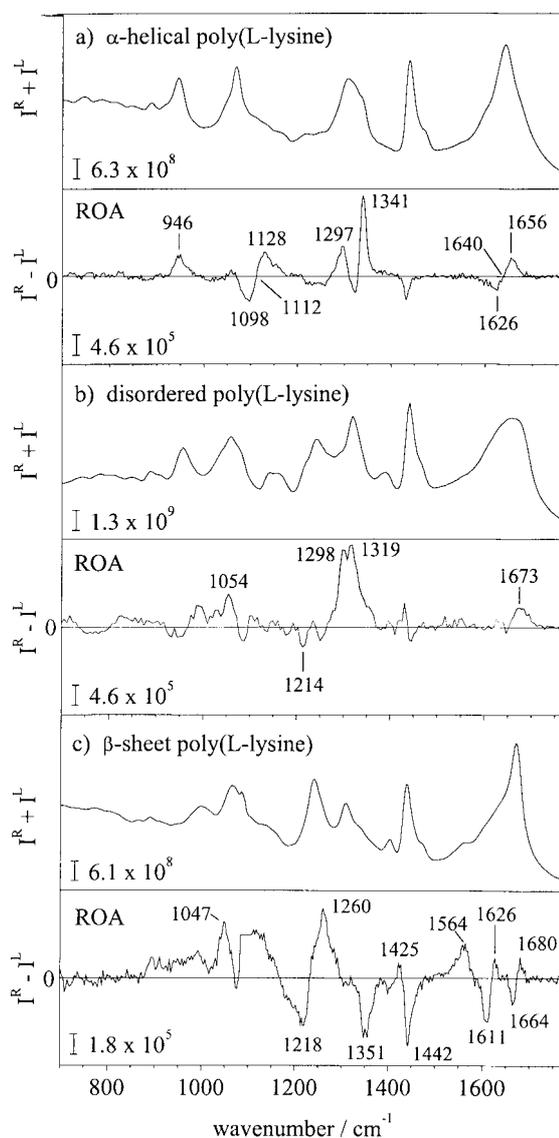


Fig. 2. Backscattered Raman ($I^R + I^L$) and ROA ($I^R - I^L$) spectra of (a) α -helical, (b) disordered, and (c) β -sheet poly(L-lysine) at 20°C.

4.2. Folded proteins

Figure 3(a) shows the backscattered Raman and ROA spectra of the α -helical protein human serum albumin. The main features of the ROA spectrum are similar to those in that of α -helical poly(L-lysine) displayed in Fig. 2(a) and accord with the Protein Data Bank (PDB) X-ray crystal structure 1ao6, which reports 69.2% α -helix, 1.7% 3_{10} -helix and the remainder being made up of loops and turns. The strong sharp positive band at $\sim 1340\text{ cm}^{-1}$ is assigned to a hydrated form of α -helix while the weaker positive band at $\sim 1300\text{ cm}^{-1}$ appears to be associated with α -helix in a more hydrophobic environment. We have reported similar bands arising from elements of α -helix in the ROA spectra of model α -helical polypeptides and other globular proteins [8,24] and the coat proteins of intact filamentous bacteriophages [27].

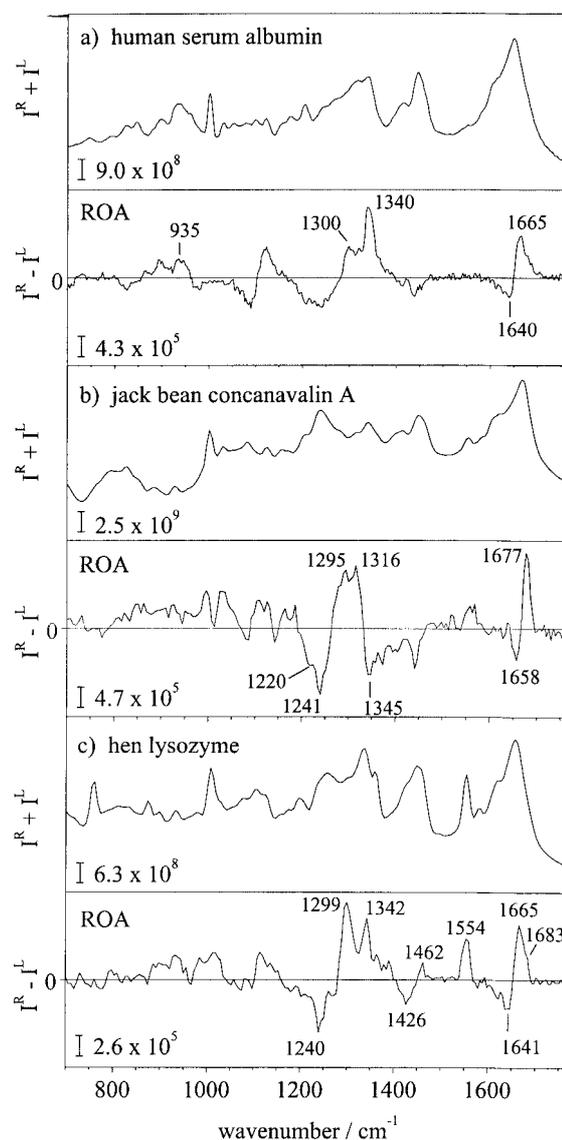


Fig. 3. Backscattered Raman and ROA spectra of (a) human serum albumin, (b) jack bean concanavalin A, and (c) hen lysozyme at pH 5.4 and 20°C.

The relative intensities of these two bands appear to correlate with the exposure of the polypeptide backbone to the solvent within the elements of α -helix in each case. For example, the positive α -helix protein ROA band at $\sim 1340 \text{ cm}^{-1}$ completely disappears when the protein is dissolved in D_2O [8]. This indicates both that the corresponding sequences are exposed to solvent and that N–H deformations of the peptide backbone make a significant contribution to the generation of this band (because corresponding N–D deformations contribute to normal modes in a spectral region several hundred wavenumbers lower). The positive ROA band from α -helix at $\sim 1300 \text{ cm}^{-1}$ retains much of its intensity in D_2O [8]. Conventional Raman bands at similar wavenumbers have been assigned to α -helix in studies performed on polypeptides [28] and filamentous bacteriophages [29]. In particular, a Raman study of α -helical poly(L-alanine)

[28] has provided definitive assignments of a number of the normal modes of vibration in the extended amide III region over the range in which we have identified ROA bands. These normal modes variously transform the same as A , E_1 and E_2 symmetry species of the point group of a model infinite regular helix. The ROA bands assigned to α -helix in this region may be related to a number of these normal modes, with the ROA intensities and exact wavenumbers being a function of the perturbations (geometric and/or due to various types of hydration) to which the particular helical sequence is subjected.

Insight into the nature of the hydrophobic and hydrated variants of α -helix has been provided by recent electron spin resonance studies of double spin-labelled alanine rich peptides which identified a new, more open conformation of the α -helix [30,31]. Computer modelling indicates that this more open geometry leaves the hydrogen bonding intact but changes the $C=O \cdots N$ angle resulting in the splaying of the backbone amide carbonyls away from the helix axis and into the solution. Therefore, the more open structure may be the preferred conformation in aqueous solution as it allows hydrogen bonding with water molecules. An equilibrium would then exist between the canonical form of α -helix, which might be responsible for the positive ROA band at $\sim 1300 \text{ cm}^{-1}$, and the more open form, which may generate the positive band at $\sim 1340 \text{ cm}^{-1}$, with the former being favoured in a hydrophobic environment and the latter in a hydrophilic (or hydrogen-bonding) environment.

The amide I ROA couplet, negative at $\sim 1640 \text{ cm}^{-1}$ and positive at $\sim 1665 \text{ cm}^{-1}$, is also characteristic of α -helix and corresponds to the wavenumber range $\sim 1645\text{--}1655 \text{ cm}^{-1}$ for α -helix bands in conventional Raman spectra [2,19,20]. Positive ROA intensity in the range $\sim 870\text{--}950 \text{ cm}^{-1}$ is a further signature of α -helix with the detailed band structure in this region appearing to show a dependence upon side chain composition, helix length and the presence of irregularities.

Figure 3(b) shows the Raman and ROA spectra of the β -sheet protein jack bean concanavalin A which contains, according to the PDB X-ray crystal structure 2cna, 43.5% β -strand, 1.7% α -helix and 1.3% 3_{10} -helix in a jelly roll β -barrel with the remainder being hairpin bends and long loops. The sharp negative band at $\sim 1241 \text{ cm}^{-1}$ has been assigned to β -structure. Similar bands have been observed in the ROA spectra of other proteins containing β -sheet [8], as sometimes has another negative band at $\sim 1220 \text{ cm}^{-1}$ which appears to be associated with a distinct variant of β -structure, possibly hydrated. As more ROA data on proteins and polypeptides containing β -sheet have accumulated it has become apparent that the true signature of β -sheet in the amide III region may be a couplet, negative at low wavenumber and positive at high. Indeed, a large conservative couplet, negative at ~ 1218 and positive at $\sim 1260 \text{ cm}^{-1}$, dominates the ROA spectrum of β -sheet poly(L-lysine) shown in Fig. 2(c). The negative peak of the couplet in proteins appears to be constrained either to the region $\sim 1220 \text{ cm}^{-1}$ or the region $\sim 1240 \text{ cm}^{-1}$. The positive peak of this couplet occurs in the range $\sim 1260\text{--}1295 \text{ cm}^{-1}$. It is possible that hydration, side chain interactions and structural irregularity may influence the positions of the negative and positive bands, and that bands from loops and turns may also contribute in this region. Thus, we now assign the positive band $\sim 1295 \text{ cm}^{-1}$ in the spectrum of concanavalin A as being the high wavenumber signal of the β -sheet couplet for this protein. Amide III bands from β -sheet in conventional Raman spectroscopy are assigned to the region $\sim 1230\text{--}1245 \text{ cm}^{-1}$ [20]. The amide I couplet, negative at $\sim 1658 \text{ cm}^{-1}$ and positive at $\sim 1677 \text{ cm}^{-1}$, is another signature of β -sheet and can be easily distinguished from the amide I couplet produced by α -helix, which typically occurs $\sim 5\text{--}20 \text{ cm}^{-1}$ lower. This correlates with β -sheet amide I bands in conventional Raman spectroscopy which occur in the range $\sim 1665\text{--}1680 \text{ cm}^{-1}$ [20]. A number of ROA bands associated with β -sheet may appear in the backbone skeletal stretch region, as found here for concanavalin A. However, the details of wavenumber, intensity and band shape are variable, possibly reflecting differences in local conformations and amino acid compositions found within different β -sheets.

Proteins containing a significant amount of β -sheet often display additional ROA signals originating in loops and turns. Negative ROA bands in the range ~ 1340 – 1380 cm^{-1} appear to originate in β -hairpin bends; an example is a band of medium intensity appearing at $\sim 1345\text{ cm}^{-1}$ in the spectrum of concanavalin A. It has been stated previously that this signature allows ROA to differentiate between parallel and antiparallel types of β -sheet as only the latter usually contain hairpin bends while the ends of strands in the former are usually connected by α -helical sequences [8]. However, this statement is only true for the parallel type of β -sheet found in α/β proteins such as those with the TIM barrel fold or the Rossman fold [32]. We have recently measured the ROA spectrum of P.69 pertactin (expressed by the virulent bacterium *Bordetella pertussis*) which has an extended β -helix fold [33] comprising β -sheet with parallel strands connected by many hairpin bends, and have observed a large negative band at $\sim 1343\text{ cm}^{-1}$ which we assign to these bends. Many β -sheet proteins also show a strong positive ROA band at ~ 1314 – 1325 cm^{-1} similar to the prominent bands observed here in disordered forms of poly(L-lysine) (Fig. 2(b)) and poly(L-glutamic acid) [8,17]. These disordered polypeptides are thought to contain significant amounts of the polyproline II (PPII) helix conformation [34,35]. This signal may therefore be a signature of the PPII helical elements known from X-ray crystal structures to occur in some of the longer loops between elements of secondary structure [36,37]. An example of such a PPII signal is observed at $\sim 1316\text{ cm}^{-1}$ in the ROA spectrum of jack bean concanavalin A.

Although the ROA spectrum of β -sheet poly(L-lysine) displayed in Fig. 2(c) contains several bands similar to those seen in concanavalin A, there are also some differences. In particular, the negative-positive-negative-positive ROA band pattern observed in the range ~ 1600 – 1690 cm^{-1} may be characteristic of extended multistranded β -sheet, the wavenumber range being similar to that observed [1] and computed [38] for the amide I vibrations of such structures. The suppressed intensity of the amide I ROA couplet centred at $\sim 1665\text{ cm}^{-1}$ relative to the large couplet seen here in concanavalin A parallels the weak amide I VCD observed for such model β -sheet structures [38] compared with the large VCD signals seen in typical β -sheet proteins. This may result from the ‘planar’ nature of the constituent strands within the relatively flat multistranded β -sheet supported by the polypeptide for which the intrinsic skeletal chirality, and hence ROA and VCD intensities, would be smaller than for the twisted strands present in the usually more irregular β -sheet structures found in typical native proteins [38]. The positive ROA band at $\sim 1561\text{ cm}^{-1}$ in β -sheet poly(L-lysine) is assigned to the amide II vibration involving an out-of-phase combination of the N–H in-plane deformation and the C_{α} –N stretch. Amide II bands are usually either weak, or are not observed at all, in the Raman and ROA spectra of proteins. We have observed that this ROA band disappears when β -sheet poly(L-lysine) is prepared in D_2O solution confirming that N–H deformations are involved, whereas the ROA in the range ~ 1600 – 1690 cm^{-1} is virtually unchanged which is consistent with its assignment to amide I vibrations. Although the ROA spectrum of β -sheet poly(L-lysine) shows a large negative band at $\sim 1351\text{ cm}^{-1}$ assigned to the hairpin bends it is expected to contain, is reassuring that there is no positive ROA in the range ~ 1314 – 1325 cm^{-1} as observed in concanavalin A and assigned to PPII-type loop structure.

Figure 3(c) shows the Raman and ROA spectra of the $\alpha + \beta$ protein hen lysozyme. The fold of this protein is very different to those of human serum albumin and concanavalin A discussed above. This is reflected in the large differences between their ROA spectra. Hen lysozyme contains 28.7% α -helix, 10.9% 3_{10} -helix and 6.2% β -sheet according to the PDB X-ray crystal structure 1lse, which is consistent with the presence of the positive ROA bands assigned to hydrophobic and hydrated α -helix at ~ 1299 and 1342 cm^{-1} , respectively, as well as the sharp negative band $\sim 1240\text{ cm}^{-1}$ indicative of β -structure. As our database of protein ROA spectra has expanded it has become apparent that the large positive peak ~ 1297 – 1300 cm^{-1} observed for proteins with a lysozyme-type fold may be boosted by bands other than

from α -helix. Proteins with a lysozyme-type fold often contain an unusually high 3_{10} -helix component. It is possible that the positive band $\sim 1299\text{ cm}^{-1}$ contains a contribution from a 3_{10} -helix band, with a positive signal at $\sim 1295\text{ cm}^{-1}$. There may also be bands from turns contributing in this region.

The amide I couplet, negative at ~ 1641 and positive at 1665 cm^{-1} with a small shoulder at $\sim 1683\text{ cm}^{-1}$, also indicates the presence of α -helix and a lesser amount of β -sheet. There is a small couplet, negative at $\sim 1426\text{ cm}^{-1}$ and positive at $\sim 1462\text{ cm}^{-1}$, from CH_2 and CH_3 side chain deformations. On the low wavenumber side of this couplet there is another relatively weak couplet, positive at low wavenumber and negative at high, that originates in tryptophan vibrations. In addition, there is a relatively large positive band $\sim 1554\text{ cm}^{-1}$ assigned to the W3-type vibrational mode of the indole ring of tryptophan residues. Miura et al. [39] found that the magnitude of the torsion angle $\chi^{2,1}$ of the tryptophan side chain, which describes the orientation of the indole ring with respect to the local peptide backbone, can be deduced from the wavenumber of the corresponding conventional Raman band. Similarly, the magnitude of the torsion angle can be determined from the position of the W3-type vibrational mode in the ROA spectrum but with the added advantage of a possible determination of its sign from the measured sign of the ROA band and hence the deduction of the absolute stereochemistry of the tryptophan side chain [40]. This information is not normally available other than from a structure determined at atomic resolution. Hen lysozyme contains six tryptophan residues, four with positive $\chi^{2,1}$ values and two with negative $\chi^{2,1}$ values according to the PDB structure 1lse. Partial cancellation of the resulting ROA signals yields the positive band observed in Fig. 3(c) [40]. The W3 ROA band can also be used as a probe of conformational heterogeneity among tryptophan residues in disordered protein sequences as the cancellation of signals with opposing signs results in a loss of ROA intensity. This is similar to the disappearance of near UVCD bands from aromatic residues upon the loss of tertiary structure [41]. These two techniques therefore provide complementary views of order/disorder transitions as ROA probes the intrinsic skeletal chirality of the tryptophan side chain while UVCD probes the chirality in the immediate vicinity of the aromatic chromophore.

4.3. Unfolded proteins

As yet it has not been possible to obtain useful ROA spectra of fully unfolded denatured states of proteins which have well-defined tertiary folds in the native state. This is because the intense Raman bands from chemical denaturants typically used at high concentration preclude ROA measurements, while thermally unfolded proteins often give rise to intense Rayleigh light scattering due to aggregation. We have, however, reported interesting results for partially unfolded denatured protein states associated with molten globules and reduced proteins, and also for proteins which are unfolded in their native biologically active states [42].

Molten globules are partially unfolded denatured protein states, stable at equilibrium, with well-defined secondary structure but lacking the specific tertiary interactions characteristic of the native state [43,44]. A much-studied molten globule is that supported by α -lactalbumins at low pH and called the A-state. Native bovine α -lactalbumin has the same fold and very similar X-ray crystal structure to that of hen lysozyme and this is apparent from its ROA spectrum [45] shown in Fig. 4(a), which is similar to that of hen lysozyme depicted in Fig. 3(c). Nevertheless, differences in detail between the spectra are apparent. This highlights the sensitivity of ROA to small differences in structure between proteins with the same fold. The Raman and ROA spectra of A-state bovine α -lactalbumin are shown in Fig. 4(b). The sensitivity of ROA to the complexity of order in molten globule states is indicated by the large differences observed here between the ROA spectra of the native and A-states of bovine α -lactalbumin as opposed to the

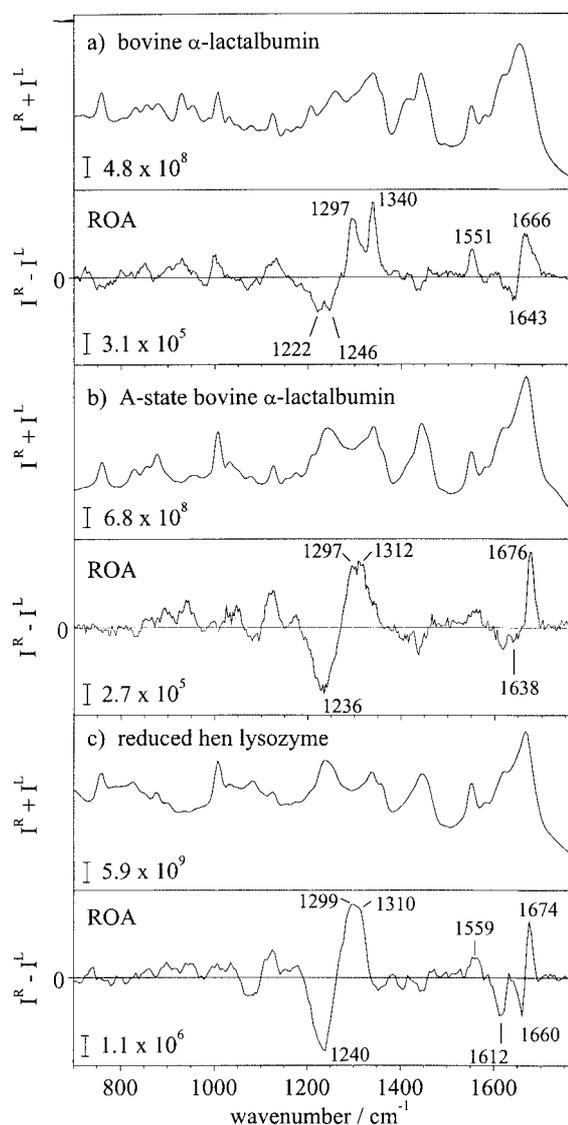


Fig. 4. Backscattered Raman and ROA spectra of (a) native bovine α -lactalbumin at pH 4.6 and 20°C, (b) A-state bovine α -lactalbumin at pH 1.9 and 2°C, and (c) reduced hen lysozyme at pH 2.4 and 20°C.

small differences between the corresponding parent Raman spectra. Much of the ROA band structure in the extended amide III region of the A-state spectrum has disappeared and is replaced by a large couplet consisting of a single negative band at $\sim 1236 \text{ cm}^{-1}$ and two distinct positive bands at ~ 1297 and 1312 cm^{-1} . The negative $\sim 1236 \text{ cm}^{-1}$ signal may originate in residues clustering around an 'average' of the conformations corresponding to the two negative signals in the native spectrum, observed at ~ 1222 and 1246 cm^{-1} in Fig. 4(a) and assigned to β -structure. The positive band at $\sim 1297 \text{ cm}^{-1}$ may arise from the same α -helical sequences in a hydrophobic environment, and possibly 3_{10} -helical sequences, as those responsible for a similar band in the spectrum of the native state. The positive $\sim 1340 \text{ cm}^{-1}$ band assigned to α -helix in a hydrated state that dominates the ROA spectrum of the native state is completely

lost in the A-state, with a new positive band appearing at $\sim 1312\text{ cm}^{-1}$ that may originate in reduced hydration in some α -helical sequences. In the amide I region the positive signal of the couplet has shifted by $\sim 10\text{ cm}^{-1}$ to higher wavenumber compared to that for the native state. The signal originating from the W3 vibrational mode of tryptophan residues at $\sim 1551\text{ cm}^{-1}$ has almost completely disappeared in the ROA spectrum of the A-state, indicating conformational heterogeneity amongst these side chains associated with a loss of the characteristic tertiary interactions found in the native state.

The Raman and ROA spectra of reduced hen lysozyme are shown in Fig. 4(c) [42]. This sample was prepared by reducing all the disulphide bonds and keeping the sample at low pH to prevent their re-oxidation. There are significant changes of the ROA spectrum of reduced lysozyme compared with that of the native state shown in Fig. 3(c). The bands in the backbone skeletal stretch region are generally suppressed, indicating the loss of much of the fixed structure. Most of the ROA band patterns in the extended amide III region have disappeared and been replaced by a large broad couplet with some hints of weak band structure, similar to that observed in the ROA spectrum of A-state α -lactalbumin, suggesting that the reduced protein supports a number of conformations with a range of local residue ϕ, ψ angles clustering in the same regions of the Ramachandran surface as in the native state. The disappearance of the positive band at $\sim 1340\text{ cm}^{-1}$ implies that none of the significant amount of hydrated α -helix found in the native state persists in the reduced form. The amide I couplet has also shifted to higher wavenumber and become quite sharp indicating the presence of a significant amount of β -structure, and a new strong negative band has appeared at $\sim 1612\text{ cm}^{-1}$. In fact the entire ROA band pattern in the range $\sim 1550\text{--}1690\text{ cm}^{-1}$ is similar to that of β -sheet poly(L-lysine) (Fig. 2(c)) suggesting the presence of multistranded sheet, perhaps arising from the presence of oligomers. The couplet, negative at ~ 1426 and positive at $\sim 1462\text{ cm}^{-1}$ in the spectrum of the native state originating in aliphatic side chains and the band at $\sim 1554\text{ cm}^{-1}$ assigned to tryptophan side chains are all greatly diminished, presumably due to conformational heterogeneity. Although of similar appearance, the ROA spectra of different denatured proteins display differences in detail, possibly reflecting the different residue compositions and their different ϕ, ψ propensities.

Although human lysozyme has a similar stability to hen lysozyme with regards to temperature and low pH, its thermal denaturation behaviour is subtly different. Below pH ~ 3.0 the thermal denaturation of human lysozyme is not a two-state process. Unlike hen lysozyme, it supports a partially folded molten globule-like state at elevated temperatures [46]. Incubation at 57°C and pH 2.0, under which conditions the partially folded state is the most highly populated, induces the formation of amyloid fibrils while incubation at 70°C and pH 2.0, under which conditions the fully denatured state is the most highly populated, leads to the formation of amorphous aggregates [47].

The Raman and ROA spectra of the native and partially folded prefibrillar amyloidogenic intermediate states of human lysozyme are shown in Figs 5(a) and 5(b), respectively [48]. The structure and ROA spectrum of the native state of human lysozyme are similar to those of hen lysozyme. However, large changes are apparent in the ROA spectrum of the prefibrillar intermediate. The most significant of these is the loss of the positive $\sim 1345\text{ cm}^{-1}$ band assigned to hydrated α -helix and the appearance of a new positive band at $\sim 1325\text{ cm}^{-1}$ assigned to PPII helix. This suggests that hydrated α -helix has undergone a conformational change to PPII structure. The disappearance of the positive $\sim 1550\text{ cm}^{-1}$ band assigned to tryptophan vibrations indicates that major conformational changes have occurred among the five tryptophan residues, four of which lie within the α -helical domain. Thus the ROA data suggest that the α -domain destabilizes and undergoes a conformational change in the prefibrillar intermediate and that PPII helix may be a critical conformational element involved in amyloid fibril formation [48]. There is no sign of an increase in β -sheet content in the intermediate. The ROA spectrum of hen lysozyme, which

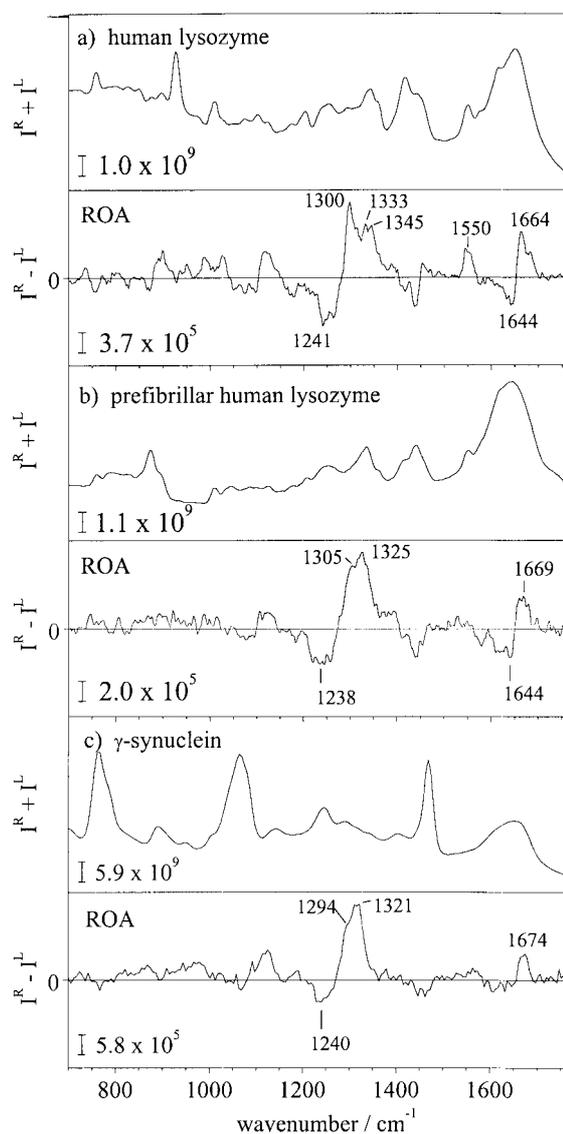


Fig. 5. Backscattered Raman and ROA spectra of native human lysozyme at pH 5.4 and 20°C, (b) the prefibrillar intermediate of human lysozyme at pH 2.0 and 57°C, and (c) human γ -synuclein at pH 7.0 and 20°C.

has a much lower propensity to form amyloid fibrils than the human variant [49], is virtually native-like under the same conditions of low pH and elevated temperature [48].

ROA has proven useful in several recent studies of natively unfolded proteins of biological and physiological significance. The ROA spectra of bovine milk caseins [50], several wheat prolamins [42], and the human recombinant synuclein and tau brain proteins [50] some of which are associated with neurodegenerative disease, were all found to be dominated by a strong positive band at $\sim 1316\text{--}1322\text{ cm}^{-1}$ assigned to PPII helix. The ROA spectrum of γ -synuclein displayed in Fig. 5(c) contains an example at $\sim 1321\text{ cm}^{-1}$. The absence of a well-defined amide I ROA couplet in this spectrum indicates the lack of secondary structure. Although the other natively unfolded proteins studied so far have ROA spectra

that are similar overall, there are differences of detail which reflect differences in residue composition and minor differences in structural elements. The studies on the caseins, synucleins and tau suggested that these proteins may be better classified as 'rheomorphic', meaning flowing shape [51], than 'random coil'. The rheomorphic state is distinct from the molten globule state which is a more compact entity containing a hydrophobic core and a significant amount of secondary structure [43,44].

The conformational plasticity supported by mobile regions within native proteins, partially denatured protein states, and natively unfolded proteins, underlies many of the conformational (protein misfolding) diseases [52,53], many of which involve amyloid fibril formation. As it is extended, flexible and hydrated, the PPII helical structure identified in some of these systems imparts a plastic open character to the structure and may be implicated in the formation of regular fibrils in the amyloid diseases [48, 50]. This is because the elimination of water molecules between extended polypeptide chains with fully hydrated N-H and C=O groups to form β -sheet hydrogen bonds is a highly favourable process entropically [54]: as the PPII- and β -sheet regions are adjacent on the Ramachandran surface it is expected that elements of PPII helix would readily undergo this type of aggregation with each other and with the edges of pre-existing β -sheet to form the cross- β structures typical of amyloid fibrils. True random coil is expected to generate the amorphous aggregates that are usually observed when most proteins are destabilized, rather than fibrils. However, although the presence of significant amounts of PPII structure may be necessary for the formation of regular fibrils, other factors may also be important since, of the milk and brain proteins studied by ROA, only α -synuclein and tau are fibrillogenic and associated with disease. It appears necessary to consider properties of the constituent residues: for example, a combination of low mean hydrophobicity and high net charge are thought to be important prerequisites for proteins to remain natively unfolded [55].

PPII helix is not readily amenable to traditional methods of structure determination due to its inherent flexibility and the lack of intrachain hydrogen bonding, which has hindered its recognition in globular proteins. For instance, the PPII conformation is indistinguishable from an irregular backbone structure for free peptides in solution by ^1H NMR spectroscopy [56]. PPII helix may be recognized in polypeptides by UVCD [34,35,56] and VCD [9,57,58], and in proteins from the deconvolution of UVCD spectra [59]. In contrast, ROA provides a clear and characteristic signature of PPII helix in both proteins and polypeptides.

The role of PPII helix in prion disease is especially interesting. An understanding of some key aspects of prion disease remain elusive, particularly the existence of an infectious form along with genetic and sporadic forms [60], something which has not been recognized in any other class of neurodegenerative disease. The basic current model involves conversion of a ubiquitous cellular form of the prion protein PrP^{C} into a scrapie (amyloid fibril) form PrP^{Sc} , the prion protein being both target and infectious agent. PrP^{C} has a predominantly α -helical structured domain together with a long disordered tail, whereas PrP^{Sc} has a high β -sheet content. A recent ROA study of full-length and truncated recombinant sheep prion protein in our laboratory has revealed that the disordered N-terminal tail of PrP^{C} is composed mainly of well-defined PPII-helical elements (E.W. Blanch, A.C. Gill, A.G.O. Rhie, J. Hope, L. Hecht and L.D. Barron, unpublished work). However, although containing a large amount of PPII structure, the disordered tail does not itself appear to be intrinsically fibrillogenic since most of the β -sheet in PrP^{Sc} appears to originate in sequences within the structured α -domain of PrP^{Sc} [60]. Indeed, the N-terminal end carries high net charge and has a low net hydrophobicity which, as mentioned above, inhibits association of separate polypeptide chains; and the disordered tail has a high proportion of proline and glycine residues which have very low β -sheet forming propensities [61]. Evidence is growing to support a functional role for PrP^{C} in copper metabolism: Cu(II) ions appear to bind to the protein in a highly

conserved octarepeat region of the N-terminal tail [62]. Presumably the rheomorphic character of the PPII structure in the N-terminal tail of the apoprotein enables the appropriate sequences to readily adapt to the fixed Cu(II)-bound conformation in the holoprotein. We have also studied a reduced form of the prion protein that takes up a β -structure at low pH. It has an ROA spectrum more similar to that of β -sheet poly(L-lysine) in Fig. 2(c) than that of typical β -sheet proteins such as concanavalin A in Fig. 3(b) suggesting the presence of β -sheet in an extended multistranded relatively flat form very different from the more irregular types of twisted β -sheet found in typical native proteins.

5. Nucleic acids

The study of the structure and function of nucleic acids remains a central problem in molecular biology. Although ROA studies of nucleic acids are at an early stage, the results obtained so far are promising as ROA appears to be sensitive to three different sources of chirality: the chiral base-stacking arrangement of intrinsically achiral base rings, the chiral disposition of the base and sugar rings with respect to the C–N glycosidic link, and the inherent chirality associated with the asymmetric centres of the sugar rings. Studies on pyrimidine nucleosides [63] and synthetic polyribonucleotides [64] have provided a basis for the interpretation of ROA spectra of DNA and RNA.

The Raman and ROA spectra of calf thymus DNA and of phenylalanine-specific transfer RNA (tRNA^{Phe}) in the presence and absence of Mg²⁺ ions are shown in Figs 6(a)–(c), respectively [65]. ROA bands in the region ~ 900 – 1150 cm⁻¹ originate in vibrations of the sugar rings and phosphate backbone. The region ~ 1200 – 1550 cm⁻¹ is dominated by normal modes in which the vibrational coordinates of the base and sugar rings are mixed. ROA band patterns in this sugar-base region appear to reflect the mutual orientation of the two rings and possibly the sugar ring conformation. The region ~ 1550 – 1750 cm⁻¹ contains ROA bands characteristic of the types of bases involved and the particular stacking arrangements. Although the ROA spectra of the DNA and the two RNAs are similar there are many differences of detail. The main differences originate in the DNA taking up a B-type double helix in which the sugar puckers are mainly C2'-endo and the RNAs taking up A-type double helical segments where the sugar puckers are mainly C3'-endo. There are smaller differences between the two RNA spectra and these are most apparent in the sugar-phosphate region ~ 900 – 1150 cm⁻¹. It is known that Mg²⁺ ions are necessary to hold RNAs in their specific tertiary folds. For tRNA^{Phe} in the presence of Mg²⁺ this is a compact L-shaped form; whereas in the absence of Mg²⁺ the tRNA^{Phe} adopts an open cloverleaf structure [66]. The ROA spectrum of the Mg²⁺-free tRNA^{Phe} shows a strong negative, positive, negative triplet at ~ 992 cm⁻¹, 1048 cm⁻¹ and 1091 cm⁻¹ which is very similar to that found in A-type polyribonucleotides and has been assigned to the C3'-endo sugar pucker. This signature is weaker and more complex in the ROA spectrum of the Mg²⁺-bound sample, suggesting a wider range of sugar puckers. Switching the sugar pucker conformations from C3'-endo to C2'-endo would elongate the sugar-phosphate backbone and may assist the formation of the loops and turns that characterize the tertiary structure of the folded form.

6. Viruses

Knowledge of the structure of viruses at the molecular level is essential for enterprises such as structure-guided antiviral drug design [67]. However, the application of key structural biology techniques

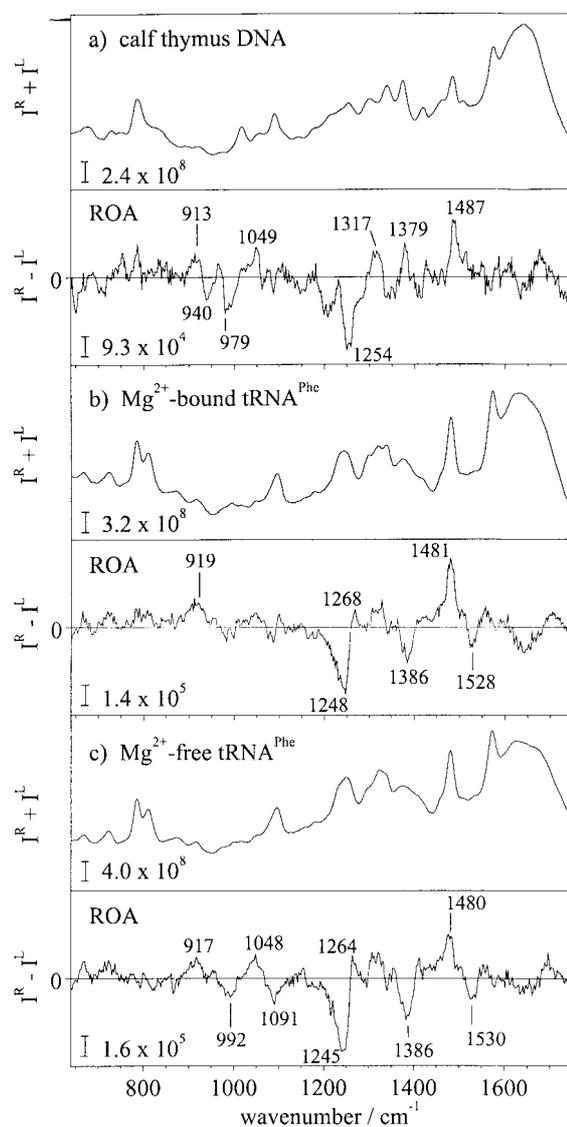


Fig. 6. Backscattered Raman and ROA spectra of (a) calf thymus DNA, (b) Mg^{2+} -bound tRNA^{Phe} , and (c) Mg^{2+} -free tRNA^{Phe} at pH 6.8 and 20°C .

such as X-ray crystallography or fibre diffraction is often hampered by practical difficulties. Conventional Raman is valuable in studies of intact viruses at the molecular level as it is able to simultaneously probe both the protein and nucleic acid constituents [20,68]. The additional incisiveness of ROA further enhances the value of Raman spectroscopy in structural virology.

The first virus ROA spectra were reported for filamentous bacteriophages [27]. The Raman and ROA spectra of three different strains, Pf1 [27], fd [69] and PH75 [40], are shown in Figs 7(a)–(c), respectively, as examples. At high concentrations (above ~ 10 – 20 mg/ml depending upon virus and detailed experimental conditions) filamentous bacteriophages form liquid crystals which give rise to artifacts which prevent the acquisition of reliable ROA data. Hence the viral concentrations were kept below the liquid

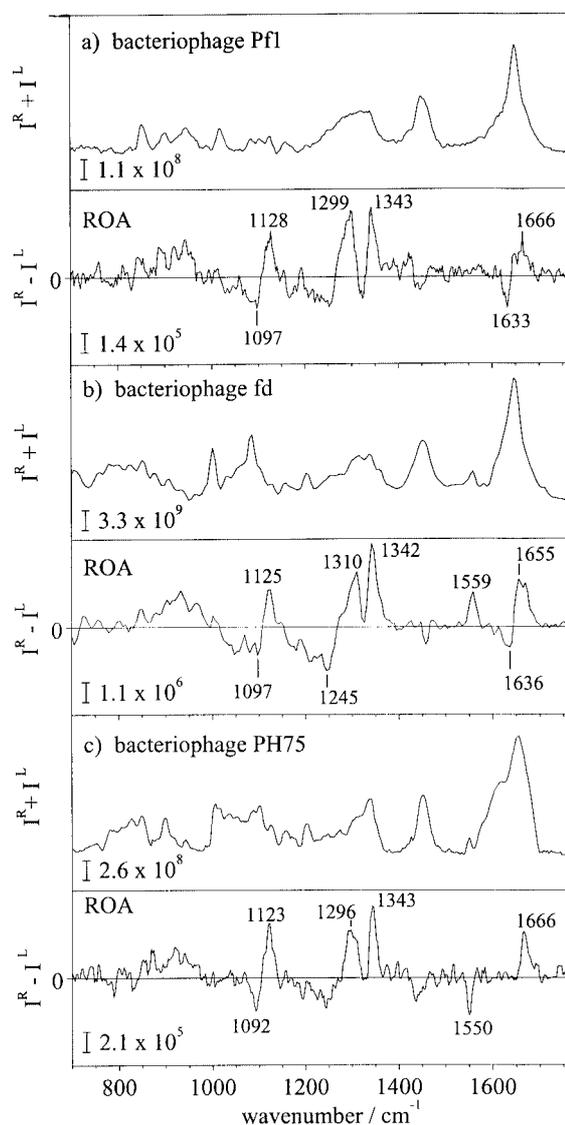


Fig. 7. Backscattered Raman and ROA spectra of the filamentous bacteriophages (a) Pf1, (b) fd, and (c) PH75 at pH 7.6 and 20°C.

crystal forming thresholds which consequently resulted in a decrease of signal-to-noise levels. This was partly offset by an apparent boost in viral protein ROA CID values, possibly due to decreased conformational mobility within closely packed environments. The water background has been subtracted from each of the corresponding parent Raman spectra shown in Fig. 7 in order to aid presentation.

Filamentous bacteriophages are long flexible rods made up of a loop of single-stranded DNA surrounded by a shell consisting of a helical array of several thousand identical copies of a major coat protein containing ~50 amino acids. There are also a few copies of minor coat proteins. X-ray fibre diffraction has shown that the major coat protein subunits have the fold of an extended α -helix and that they overlap each other so that the exterior half of each protein is exposed to water while the interior half is protected by neighbouring coat proteins [70]. The ROA spectra are dominated by bands assigned to α -helix in the

backbone skeletal stretch, amide I and extended amide III regions which have been discussed already. Further, the relative intensities of the ROA bands at ~ 1300 and 1342 cm^{-1} assigned to hydrophobic and hydrated α -helix, respectively, correlate well with the hydration of the peptide backbone in each coat protein [27]. The interior protected region of the coat protein of Pf1 contains a continuous sequence of 20 hydrophobic residues while the exterior exposed region contains a 20-residue sequence of mixed hydrophobic and hydrophilic residues. In Fig. 7(a) the intensities of the corresponding ROA bands at ~ 1299 and 1343 cm^{-1} are virtually identical. For bacteriophage fd the sequence of pure hydrophobic helix in the interior region of the coat protein is shorter and there is a longer sequence of mixed hydrophobic and hydrophilic residues. In the corresponding ROA spectrum in Fig. 7(b) the hydrated α -helix band is now relatively stronger than the band originating in hydrophobic α -helix.

An obvious difference between the three bacteriophage ROA spectra is the behaviour of the band associated with the W3-type tryptophan side-chain vibration. The major coat protein of Pf1 does not contain any tryptophan residues and there is no ROA signal at $\sim 1550\text{ cm}^{-1}$. Bacteriophage fd contains a single tryptophan residue in the major coat protein which generates the *positive* ROA band at $\sim 1559\text{ cm}^{-1}$ while PH75, which has unusual thermophilic properties [71], shows a *negative* band at $\sim 1550\text{ cm}^{-1}$ originating from the single tryptophan in its coat protein. This suggests that the absolute stereochemistry of the indole ring of the tryptophan residue relative to the local peptide backbone in fd is quasi-enantiomeric to that in PH75. From the wavenumbers and signs of the W3 ROA bands it was deduced that $\chi^{2,1} = +120^\circ$ for fd [69] and -93° for PH75 [40].

Figures 8(a)–(c) show, respectively, the Raman and ROA spectra of tobacco mosaic virus (TMV), potato virus X (PVX) and narcissus mosaic virus (NMV) [72]. These are helical plant viruses containing a single strand of positive-sense RNA encapsidated within a rigid rod-shaped particle in the case of TMV, or a flexuous filamentous particle in the case of PVX and NMV, made up of multiple copies of a single coat protein. The structure of the coat protein subunits of TMV has been determined by X-ray fibre diffraction [73] to be based on a four-helix bundle motif which contains both water-exposed residues and residues at the hydrophobic interfaces between α -helices. The ROA band pattern in the extended amide III region of the TMV spectrum is characteristic of a helix bundle, with a positive band at $\sim 1295\text{ cm}^{-1}$ assigned to α -helix in a hydrophobic environment and a slightly stronger band at $\sim 1342\text{ cm}^{-1}$ assigned to hydrated α -helix. The amide I couplet and bands in the backbone skeletal stretch region are also characteristic of α -helix.

Little is known about the conformations of the coat protein subunits of PVX and NMV from other techniques but the similarities of their ROA spectra to the ROA spectrum of TMV reveal they are both based on helix bundle folds. All three ROA spectra also contain a positive band at $\sim 1316\text{ cm}^{-1}$ assigned to elements of PPII helix in long loops and a small negative band at $\sim 1220\text{ cm}^{-1}$ from segments of β -strand, possibly hydrated. Although the three ROA spectra are similar, the differences in detail illustrate the sensitivity of ROA to different structural parameters and types of subunit packing within different virus capsids.

In addition to filamentous and rod-shaped viruses, ROA has also been applied to icosahedral viruses [69]. Satellite tobacco mosaic virus (STMV) is a small T=1 icosahedral virus with a single strand of RNA inside a capsid composed of 60 identical copies of a coat protein with 159 amino acids. The X-ray crystal structure of STMV [74] reveals the fold of the coat protein to be a jelly roll β -barrel. Although the overall appearance of the ROA spectrum of STMV, shown in Fig. 9(a), is similar to that of proteins with a jelly roll fold such as concanavalin A (Fig. 3(b)), there are some small differences due to contributions from RNA bands. Bacteriophage MS2 is a small T=3 icosahedral virus with a single strand of RNA contained inside a capsid made up of 180 identical copies of a coat protein containing 129 amino acids.

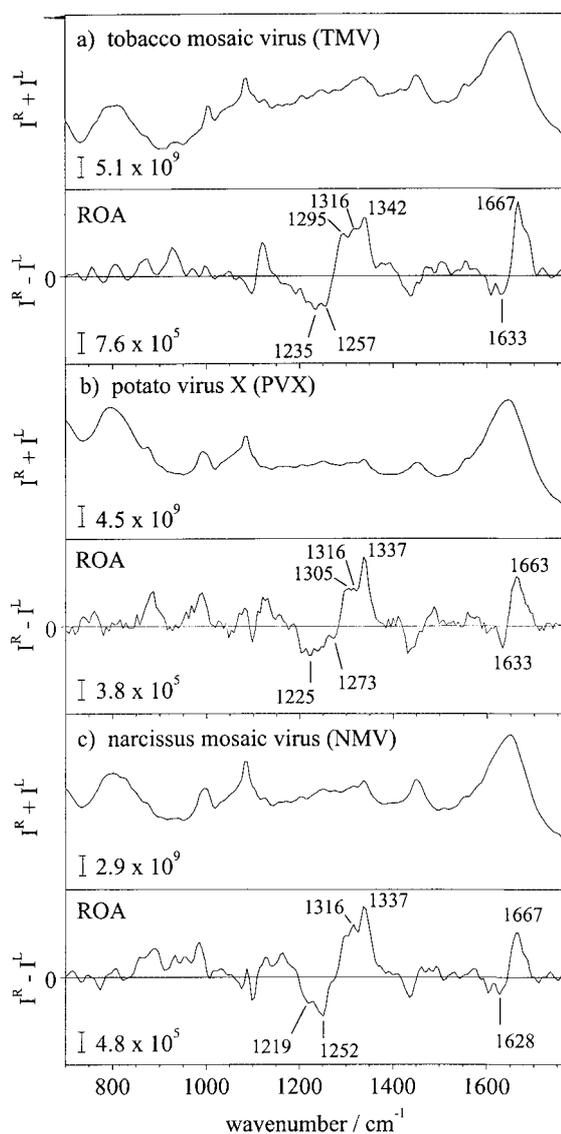


Fig. 8. Backscattered Raman and ROA spectra of the helical plant viruses (a) TMV, (b) PVX, and (c) NMV at pH 7.4 and 20°C.

The X-ray crystal structure of MS2 [75] reveals that the coat protein subunits fold into a five-stranded antiparallel β -sheet with an additional short hairpin at the N-terminus and two α -helices. The α -helices are responsible for interactions with a second subunit to form a dimer containing 10 adjacent antiparallel β -strands. This coat protein fold is quite different from the jelly roll type found in STMV and most other icosahedral viruses. The ROA spectrum of the empty MS2 protein capsid is shown in Fig. 9(b). Although clearly characteristic of a structure rich in β -sheet, there are some differences of detail compared with the ROA spectrum of STMV which reflect the two distinct types of β -sheet folds.

Determination of the structures of the nucleic acid component of intact viruses has proven difficult in even the best-resolved X-ray crystal structures. Several of the ROA spectra of viruses recorded in our laboratory have displayed bands attributed to nucleic acid. However, these are usually weak because of

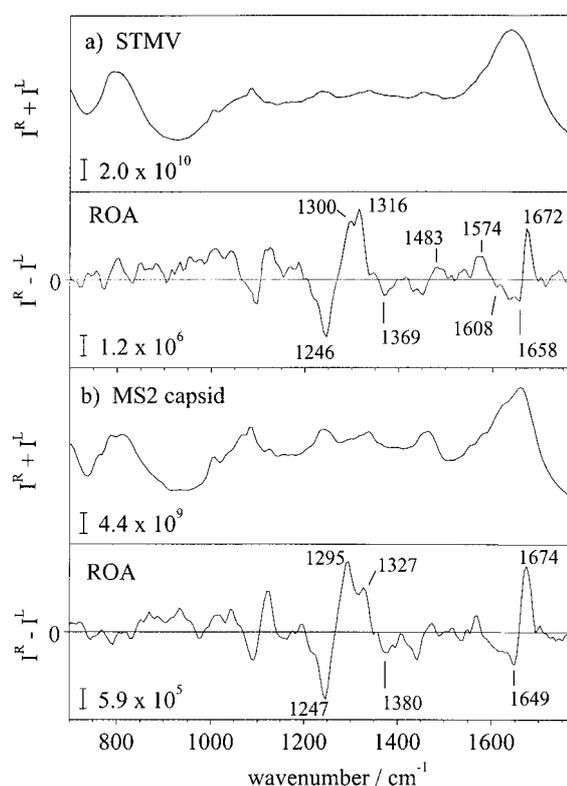


Fig. 9. Backscattered Raman and ROA spectra of the icosahedral viruses (a) STMV in H₂O at 4°C and (b) MS2 protein capsid at pH 7.2 and 20°C.

either the low nucleic acid content of the virus or its flexible nature. Recently, though, we have recorded the first nucleic acid ROA spectra from an intact virus [69], namely cowpea mosaic virus (CPMV) which is the type member of the comovirus genus. CPMV has a bipartite genome, meaning that two RNA molecules (called RNA-1 and RNA-2) are separately encapsidated. CPMV particles can be separated into four distinct isolates by ultracentrifugation with a CsCl gradient [76]. The top component (T-CPMV) consists of empty protein capsids, the middle component (M-CPMV) contains capsids with a molecule of RNA-2 bound, the bottom-upper component (B_U-CPMV) contains capsids with RNA-1 bound, and the bottom-lower component (B_L-CPMV) contains capsids with RNA-2 bound plus Cs⁺ ions permeating the interior of the particles. The Raman and ROA spectra of T-CPMV and M-CPMV are shown in Figs 10(a) and 10(b), respectively. CPMV has an icosahedral capsid of T=3 symmetry, with two protein chains forming a large (L) coat protein and a third protein chain forming a small (S) coat protein. Despite the low sequence homology between these three protein chains, X-ray crystallography shows that they exhibit very similar eight-stranded jelly roll β -barrel folds [77]. The details of the T-CPMV ROA spectrum are consistent with those of a jelly roll β -barrel fold (see the ROA spectrum of concanavalin A in Fig. 3(b)). There are several large changes in the ROA spectrum of M-CPMV due to new bands originating from the RNA-2 molecule, which constitutes ~24% of the particle by mass. These RNA bands are isolated in Fig. 10(c) which shows the Raman and ROA spectra of the bound RNA-2 molecule obtained by the subtraction (M-CPMV)-(T-CPMV). Bands corresponding to conformational changes induced in the protein structure by nucleic acid binding may also be present. Comparison of this ROA spectrum with

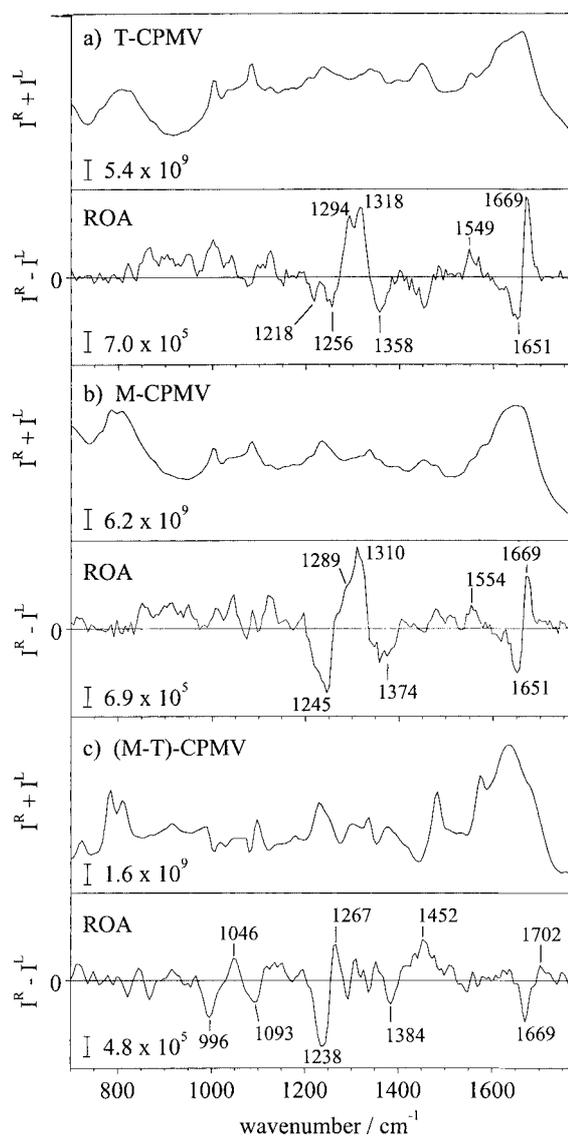


Fig. 10. Backscattered Raman and ROA spectra of the icosahedral viruses (a) T-CPMV and (b) M-CPMV at pH 7.0 and 20°C, and (c) the difference (b) – (a) (M-CPMV) – (T-CPMV).

those of the nucleic acids shown in Fig. 6 indicates that RNA-2 has an A-type helical conformation. A similar analysis has also shown that the RNA-1 molecule in the B_U-CPMV particle has the same conformation. This work has provided new information on the RNA structure of CPMV since the nucleic acid is not observed in the X-ray crystal structure [77].

7. Principal component analysis

Since protein ROA spectra contain bands characteristic of loops and turns in addition to bands characteristic of secondary structure, they should provide information about the overall three-dimensional solu-

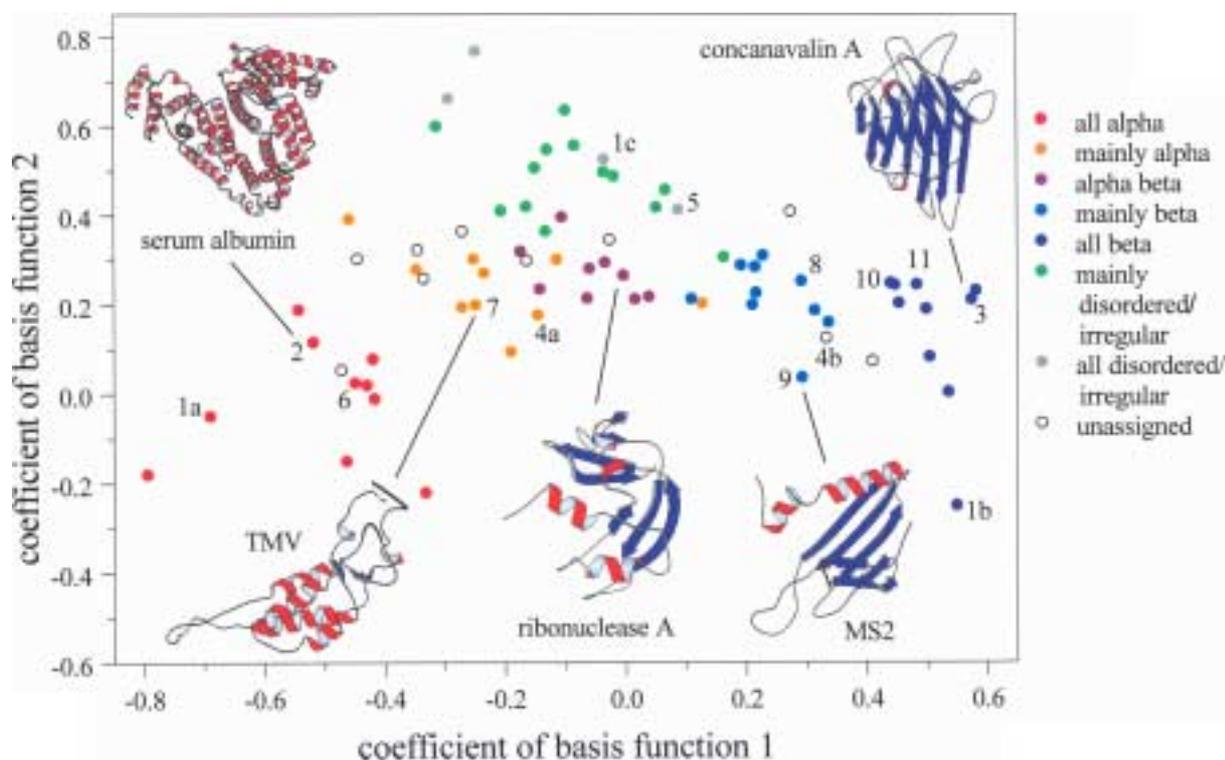


Fig. 11. Plot of the PCA coefficients for the two most important basis functions for a set of 75 polypeptide, protein and virus ROA spectra. Some of the examples discussed in this article are labelled as follows: α -helical (1a), β -sheet (1b) and disordered (1c) poly(L-lysine); human serum albumin (2); jack bean concanavalin A (3); native (4a) and reduced (4b) hen lysozyme; human γ -synuclein (5); bacteriophage fd (6); TMV (7); STMV (8), MS2 capsid (9), T-CPMV (10) and M-CPMV (11). More complete definitions of the structural types are as follows: all alpha, $>\sim 60\%$ α -helix with little or no other secondary structure; mainly alpha, $>\sim 35\%$ α -helix and a small amount of β -sheet ($\sim 5\text{--}15\%$); alpha beta, significant amounts of α -helix and β -sheet; mainly beta, $>\sim 35\%$ β -sheet and a small amount of α -helix ($\sim 5\text{--}15\%$); all beta, $>\sim 45\%$ β -sheet with little or no other secondary structure; mainly disordered/irregular, little secondary structure; all disordered/irregular, no secondary structure. MOLSCRIPT diagrams of examples of the all alpha (human serum albumin, PDB code 1ao6), mainly alpha (TMV coat protein, PDB code 1vtm), alpha beta (bovine ribonuclease A, PDB code 1rbx), mainly beta (MS2 coat protein, PDB code 1msc) and all beta (jack bean concanavalin A, PDB code 2cna) fold types are shown.

tion structure. We are developing a pattern recognition program, based on principal component analysis (PCA), to identify protein folds from ROA spectral band patterns [72]. The method is similar to one developed for the determination of the structure of proteins from VCD [78] and UVCD [79] spectra, but is expected to provide enhanced discrimination between different structural types since protein ROA spectra contain many more structure-sensitive bands than either VCD or UVCD. From the ROA spectral data, the PCA program 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. Our current set contains 75 entries comprising ROA spectra of polypeptides in model α -helical, β -sheet and disordered states, many proteins with well-defined tertiary folds known mostly from X-ray crystallography with a few from multidimensional solution NMR, and several viruses with coat protein folds known from X-ray crystallography or fibre diffraction. The method appears to be useful for unfolded and partially folded as well as completely folded protein structures.

Figure 11 shows a plot of the coefficients for the whole set of ROA spectra for the two most important raw basis functions. This serves to separate the spectra into clusters corresponding to different types of protein structure, which enables structural similarities between proteins of unknown structure with those of known structure to be identified. The polypeptides, proteins and viruses are colour coded with respect to the seven different structural types listed on the figure, which provide a useful initial classification that will be refined in later work to provide, among other things, quantitative estimates of the various types of structural elements such as helix, sheet, loops and turns. MOLSCRIPT diagrams [80] are provided of examples of the five well-defined fold types. The plot reveals increasing α -helix content to the left, increasing β -sheet content to the right, and increasing disordered or irregular structure from bottom to top. The positions of a few of the polypeptides, proteins and viruses discussed in this article are numbered according to the list provided in the figure caption.

The α -helical, β -sheet and disordered states of poly(L-lysine) all appear in the correct regions of Fig. 11, as do the native proteins human serum albumin, jack bean concanavalin A and hen lysozyme. The position of reduced hen lysozyme in the mainly β -region, well to the right of that of the native protein, suggests that it contains significantly more β -sheet than the native protein, in accordance with the analysis of its ROA spectrum given earlier. Human γ -synuclein lies close to disordered poly(L-lysine) a little way down from the top suggesting that it is composed mainly but not completely of disordered or irregular structure, as expected for a natively unfolded protein. Bacteriophage fd and TMV fall correctly within the all alpha and mainly alpha regions, respectively, in accordance with their known coat protein subunit structures. T- and M-CPMV fall on the left side of the all beta region, consistent with their jelly roll fold. Although the presence of nucleic acid bands in the ROA spectra of intact viruses will affect their positions in the plot, this is not likely to be too significant because basis functions 1 and 2 pertain mostly to protein structure: nucleic acid bands contribute more to the higher-order basis functions. This is borne out by the fact that T- and M-CPMV fall closely together in the plot despite the ROA spectra of M-CPMV containing strong RNA bands that are absent from that of T-CPMV. With planned developments of our PCA program, it should be possible to use PCA to separate the ROA spectrum of an intact virus into its protein and nucleic acid components, thereby eliminating the need for the empty protein capsid.

8. Conclusion

The sensitivity of ROA to molecular chirality makes it a valuable new tool for investigating biomolecular structure and behaviour in solution. ROA may now be applied routinely to a wide range of proteins, nucleic acids and viruses and is already providing information complementary to that obtained from high resolution techniques such as X-ray crystallography and NMR spectroscopy. The large number of resolved structure-sensitive bands in protein ROA spectra compared with other spectroscopic techniques makes pattern recognition methods such as PCA especially valuable. This will greatly facilitate applications of ROA to high throughput protein fold recognition in structural proteomics, to the protein misfolding diseases and to structural virology. With the expected availability of a commercial instrument in the near future, we hope that this brief review will encourage wide use of vibrational ROA spectroscopy in biomolecular science.

Acknowledgements

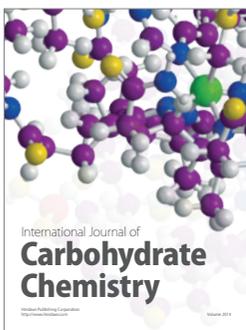
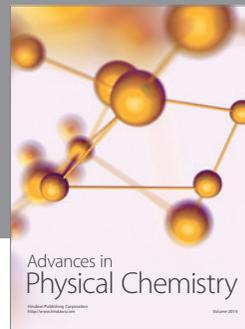
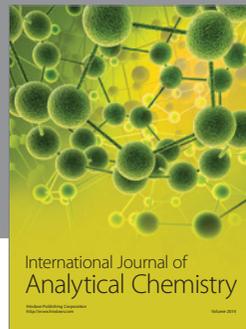
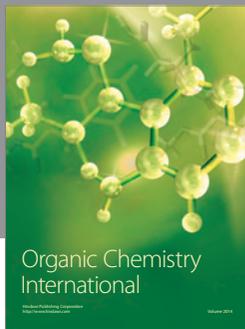
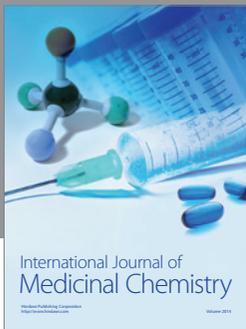
We thank the EPSRC and BBSRC for research grants, and the students and collaborators who have contributed to the Glasgow biomolecular ROA programme over many years.

References

- [1] B.R. Singh, ed., *Infrared Analysis of Peptides and Proteins: Principles and Applications*, ACS Symposium Series 750, American Chemical Society, Washington, DC, 2000.
- [2] P.R. Carey, *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York, 1982.
- [3] M. Diem, *Modern Vibrational Spectroscopy*, Wiley, New York, 1993.
- [4] P.L. Polavarapu, *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*, Elsevier, Amsterdam, 1998.
- [5] L.D. Barron and L. Hecht, Vibrational Raman optical activity: from fundamentals to biochemical applications, in: *Circular Dichroism. Principles and Applications*, 2nd edn, N. Berova, K. Nakanishi and R.W. Woody, eds, Wiley, New York, 2000, pp. 667–701.
- [6] L.A. Nafie, Infrared and Raman vibrational optical activity, *Ann. Rev. Phys. Chem.* **48** (1997), 357–386.
- [7] L.D. Barron, M.P. Bogaard and A.D. Buckingham, Raman scattering of circularly polarized light by optically active molecules, *J. Am. Chem. Soc.* **95** (1973), 603–605.
- [8] L.D. Barron, L. Hecht, E.W. Blanch and A.F. Bell, Solution structure and dynamics of biomolecules from Raman optical activity, *Prog. Biophys. Mol. Biol.* **73** (2000), 1–49.
- [9] T.A. Keiderling, Peptide and protein conformational studies with vibrational circular dichroism and related spectroscopies, in: *Circular Dichroism: Principles and Applications*, 2nd edn, N. Berova, K. Nakanishi and R.W. Woody, eds, Wiley, New York, 2000, pp. 621–666.
- [10] P.W. Atkins and L.D. Barron, Rayleigh scattering of polarized photons by molecules, *Mol. Phys.* **16** (1969), 453–466.
- [11] L.D. Barron and A.D. Buckingham, Rayleigh and Raman scattering from optically active molecules, *Mol. Phys.* **20** (1971), 1111–1119.
- [12] A.D. Buckingham, Permanent and induced molecular moments and long-range intermolecular forces, *Adv. Chem. Phys.* **12** (1967), 107–142.
- [13] L.D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
- [14] L.D. Barron and A.D. Buckingham, A simple two-group model for Rayleigh and Raman optical activity, *J. Am. Chem. Soc.* **96** (1974), 4769–4773.
- [15] L. Hecht, L.D. Barron and W. Hug, Vibrational Raman optical activity in backscattering, *Chem. Phys. Lett.* **158** (1989), 341–344.
- [16] L. Hecht and L.D. Barron, Recent developments in Raman optical activity instrumentation, *Faraday Discuss.* **99** (1994), 35–47.
- [17] L. Hecht, L.D. Barron, E.W. Blanch, A.F. Bell and L.A. Day, Raman optical activity instrument for studies of biopolymer structure and function, *J. Raman Spectrosc.* **30** (1999), 815–825.
- [18] W. Hug and G. Hangartner, A novel high-throughput Raman spectrometer for polarization difference measurements, *J. Raman Spectrosc.* **30** (1999), 841–852.
- [19] A.T. Tu, Peptide backbone conformation and microenvironment of protein sidechains, *Adv. Spectrosc.* **13** (1986), 47–112.
- [20] T. Miura and G.J. Thomas, Jr., Raman spectroscopy of proteins and their assemblies, in: *Subcellular Biochemistry*, Vol. 24. *Proteins: Structure, Function and Engineering*, B.B. Biswas and S. Roy, eds, Plenum Press, New York, 1995, pp. 55–99.
- [21] S.J. Ford, Z.Q. Wen, L. Hecht and L.D. Barron, Vibrational Raman optical activity of alanyl peptide oligomers: a new perspective on aqueous solution conformation, *Biopolymers* **34** (1994), 303–313.
- [22] R. Schweitzer-Stenner, Visible and uv-resonance Raman spectroscopy of model peptides, *J. Raman Spectrosc.* **32** (2001), 711–732.
- [23] S.A. Overman and G.J. Thomas, Jr., Raman markers of nonaromatic side chains in an α -helix assembly: Ala, Asp, Glu, Gly, Ile, Leu, Lys, Ser and Val residues of phage fd subunits, *Biochemistry* **38** (1999), 4018–4027.
- [24] G. Wilson, L. Hecht and L.D. Barron, Vibrational Raman optical activity of α -helical and unordered poly(L-lysine), *J. Chem. Soc. Faraday Trans.* **92** (1996), 1503–1510.
- [25] B. Davidson and G.D. Fasman, The conformational transitions of uncharged poly-L-lysine. α helix–random coil– β structure, *Biochemistry* **6** (1967), 1616–1629.
- [26] L.D. Barron, L. Hecht, A.F. Bell and G. Wilson, Recent developments in Raman optical activity of biopolymers, *Appl. Spectrosc.* **50** (1996), 619–629.
- [27] E.W. Blanch, A.F. Bell, L. Hecht, L.A. Day and L.D. Barron, Raman optical activity of filamentous bacteriophages: hydration of α -helices, *J. Mol. Biol.* **290** (1999), 1–7.
- [28] S.-H. Lee and S. Krimm, Polarized Raman spectra of oriented films of α -helical poly(L-alanine) and its N-deuterated analogue, *J. Raman Spectrosc.* **29** (1998), 73–80.
- [29] M. Tsuboi, M. Suzuki, S.A. Overman and G.J. Thomas, Jr., Intensity of the polarized Raman band at 1340–1345 cm^{-1} as an indicator of protein α -helix orientation: application to Pf1 filamentous virus, *Biochemistry* **39** (2000), 2677–2684.
- [30] K.A. Bolin and G.L. Millhauser, α and 3_{10} : the split personality of polypeptide helices, *Accs. Chem. Res.* **32** (1999), 1027–1033.

- [31] P. Hanson, D.J. Anderson, G. Martinez, G. Millhauser, F. Formaggio, M. Crisma, C. Toniolo and C. Vita, Electron spin resonance and structural analysis of water soluble, alanine-rich peptides incorporating TOAC, *Mol. Phys.* **95** (1998), 957–966.
- [32] C. Branden and J. Tooze, *Introduction to Protein Structure*, 2nd edn, Garland Publishing, New York, 1999.
- [33] P. Emsley, I.G. Charles, N.F. Fairweather and N.W. Isaacs, Structure of *Bordetella pertussis* virulence factor P.69 pertactin, *Nature* **381**, 90–92.
- [34] M.L. Tiffany and S. Krimm, New chain conformations of poly(glutamic acid) and polylysine, *Biopolymers* **6** (1968), 1379–1382.
- [35] R.W. Woody, Circular dichroism and conformations of unordered polypeptides, *Adv. Biophys. Chem.* **2** (1992), 37–79.
- [36] A.A. Adzhubei and M.J.E. Sternberg, Left-handed polyproline II helices commonly occur in globular proteins, *J. Mol. Biol.* **229** (1993), 472–493.
- [37] B.T. Stapley and T.P. Creamer, A survey of left-handed polyproline II helices, *Protein Sci.* **8** (1999), 587–595.
- [38] J. Kubelka and T.A. Keiderling, Differentiation of β -sheet forming structures: Ab initio-based simulations of IR absorption and vibrational CD for model peptide and protein β -sheets, *J. Am. Chem. Soc.* **123** (2001), 12 048–12 058.
- [39] T. Miura, H. Takeuchi and I. Harada, Tryptophan Raman bands sensitive to hydrogen bonding and side-chain conformation, *J. Raman Spectrosc.* **20** (1989), 667–671.
- [40] E.W. Blanch, L. Hecht, L.A. Day, D.M. Pederson and L.D. Barron, Tryptophan absolute stereochemistry in viral coat proteins from Raman optical activity, *J. Am. Chem. Soc.* **123** (2001), 4863–4864.
- [41] R.W. Woody and A.K. Dunker, Aromatic and cystine side-chain circular dichroism in proteins, in: *Circular Dichroism and the Conformational Analysis of Biomolecules*, G.D. Fasman, ed., Plenum Press, New York, 1996, pp. 109–157.
- [42] L.D. Barron, E.W. Blanch and L. Hecht, Unfolded proteins studied by Raman optical activity, *Adv. Prot. Chem.*, in press.
- [43] O.B. Ptitsyn, Molten globule and protein folding, *Adv. Prot. Chem.* **47** (1995), 83–229.
- [44] M. Arai and K. Kuwajima, The role of the molten globule state in protein folding, *Adv. Prot. Chem.* **53** (2000), 209–282.
- [45] E.W. Blanch, L.A. Morozova-Roche, L. Hecht, W. Noppe and L.D. Barron, Raman optical activity characterization of native and molten globule states of equine lysozyme: comparison with hen lysozyme and bovine α -lactalbumin, *Biopolymers* **57** (2000), 235–248.
- [46] P. Haezebrouck, M. Joniau, H. van Dael, S.D. Hooke, N.D. Woodruff and C.M. Dobson, An equilibrium partially folded state of human lysozyme at low pH, *J. Mol. Biol.* **246** (1995), 382–387.
- [47] L.A. Morozova-Roche, J. Zurdo, A. Spencer, W. Noppe, V. Receveur, D.B. Archer, M. Joniau and C.M. Dobson, Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants, *J. Struct. Biol.* **130** (2000), 339–351.
- [48] E.W. Blanch, L.A. Morozova-Roche, D.A.E. Cochran, A.J. Doig, L. Hecht and L.D. Barron, Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme, *J. Mol. Biol.* **301** (2000), 553–563.
- [49] M.R.H. Krebs, D.K. Wilkins, E.W. Chung, M.C. Pitkeathly, A.K. Chamberlain, J. Zurdo, C.V. Robinson and C.M. Dobson, Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the β -domain, *J. Mol. Biol.* **300** (2000), 541–549.
- [50] C.D. Syme, E.W. Blanch, C. Holt, R. Jakes, M. Goedert, L. Hecht and L.D. Barron, A Raman optical activity study of rheomorphism in caseins, synucleins and tau: New insight into the structure and behaviour of natively unfolded proteins, *Eur. J. Biochem.* **269** (2002), 148–156.
- [51] C. Holt and L. Sawyer, Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the α_{S1} -, β - and κ -caseins, *J. Chem. Soc. Faraday Trans.* **89** (1993), 2683–2692.
- [52] R.W. Carrell and D.A. Lomas, Conformational disease, *The Lancet* **350** (1997), 134–138.
- [53] C.M. Dobson, R.J. Ellis and A.R. Fersht, eds, *Protein Misfolding and Disease*, *Philos. Trans. R. Soc. Lond. B* **356** (2001), 127–227.
- [54] Y.C. Sekharudu and M. Sundaralingham, Hydration of protein secondary structures – the role in protein folding, in: *Water and Biological Macromolecules*, E. Westhof, ed., CRC Press, Boca Raton, FL, 1993, pp. 148–162.
- [55] V.N. Uversky, J.R. Gillespie and A.L. Fink, Why are ‘natively unfolded’ proteins unstructured under physiologic conditions, *Proteins* **41** (2000), 415–427.
- [56] G. Siligardi and A.F. Drake, The importance of extended conformations and, in particular, the P_{II} conformation for the molecular recognition of peptides, *Biopolymers* **37** (1995), 281–292.
- [57] R.K. Dukor and T.A. Keiderling, Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides, *Biopolymers* **31** (1991), 1747–1761.
- [58] T.A. Keiderling, R.A.G.D. Silva, G. Yoder and R.K. Dukor, Vibrational circular dichroism of selected oligopeptide conformations, *Bioorg. Med. Chem.* **7** (1999), 133–141.
- [59] N. Sreerama and R.W. Woody, Poly(pro)II helices in globular proteins: identification and circular dichroic analysis, *Biochemistry* **33** (1994), 10 022–10 025.
- [60] S.B. Prusiner, Prions, *Proc. Natl. Acad. Sci. USA* **95** (1998), 13 363–13 383.

- [61] C.K. Smith and L. Regan, Construction and design of β -sheets, *Accs. Chem. Res.* **30** (1997), 153–161.
- [62] J.H. Viles, F.E. Cohen, S.B. Prusiner, D.B. Goodin, P.E. Wright and H.J. Dyson, Copper binding to the prion protein: Structural implications of four identical cooperative binding sites, *Proc. Natl. Acad. Sci. USA* **96** (1999), 2042–2047.
- [63] A.F. Bell, L. Hecht and L.D. Barron, Vibrational Raman optical activity of pyrimidine nucleosides, *J. Chem. Soc. Faraday Trans.* **93** (1997), 553–562.
- [64] A.F. Bell, L. Hecht and L.D. Barron, Vibrational Raman optical activity as a probe of polyribonucleotide solution stereochemistry, *J. Am. Chem. Soc.* **119** (1997), 6006–6013.
- [65] A.F. Bell, L. Hecht and L.D. Barron, Vibrational Raman optical activity of DNA and RNA, *J. Am. Chem. Soc.* **120** (1998), 5820–5821.
- [66] W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
- [67] W. Chiu, R.M. Burnett and R.L. Garcia, eds, *Structural Virology of Viruses*, Oxford University Press, New York, 1997.
- [68] G.J. Thomas, Jr., Raman spectroscopy of protein and nucleic acid assemblies, *Ann. Rev. Biophys. Biomol. Struct.* **28** (1999), 1–27.
- [69] E.W. Blanch, L. Hecht, C.D. Syme, V. Volpetti, G.P. Lomonosoff and L.D. Barron, Molecular structure of viruses from Raman optical activity, *J. Gen. Virol.*, in press.
- [70] D.A. Marvin, Filamentous phage structure, infection and assembly, *Curr. Opin. Struct. Biol.* **8** (1998), 150–158.
- [71] D.M. Pederson, L.C. Welsh, D.A. Marvin, M. Sampson, R.N. Perham, M. Yu and M.R. Slater, The protein capsid of filamentous bacteriophage PH75 from *thermus thermophilus*, *J. Mol. Biol.* **309** (2001), 401–421.
- [72] E.W. Blanch, D.J. Robinson, L. Hecht, C.D. Syme, K. Nielsen and L.D. Barron, Solution structures of potato virus X and narcissus mosaic virus from Raman optical activity, *J. Gen. Virol.* **83** (2002), 241–246.
- [73] K. Namba, R. Pattanayek and G. Stubbs, Visualization of protein–nucleic acid interactions in a virus, *J. Mol. Biol.* **208** (1989), 307–325.
- [74] S.B. Larson, J. Day, A. Greenwood and A. McPherson, Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution, *J. Mol. Biol.* **277** (1998), 37–59.
- [75] K. Valegård, L. Liljas, K. Fridborg and T. Unge, The three-dimensional structure of the bacterial virus MS2, *Nature* **345** (1990), 36–41.
- [76] G.P. Lomonosoff and J.E. Johnson, The synthesis and structure of comovirus capsids, *Prog. Biophys. Mol. Biol.* **55** (1991), 107–137.
- [77] T. Lin, Z. Chen, R. Usha, C.V. Stauffacher, J.-B. Dai, T. Schmidt and J.E. Johnson, The refined crystal structure of cowpea mosaic virus at 2.8 Å resolution, *Virology* **265** (1999), 20–34.
- [78] P. Pancoska, S.C. Yasui and T.A. Keiderling, Statistical analysis of the vibrational circular dichroism of selected proteins and relationship to secondary structures, *Biochemistry* **30** (1991), 5089–5103.
- [79] S.Y. Venyaminov and J.T. Yang, Determination of protein secondary structure, in: *Circular Dichroism and the Conformational Analysis of Biomolecules*, G.D. Fasman, ed., Plenum Press, New York, 1996, pp. 69–107.
- [80] P.J. Kraulis, MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Cryst.* **24** (1991), 946–950.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

