Molecular structures of viruses from Raman optical activity

Ewan W. Blanch,1 Lutz Hecht,1 Christopher D. Syme,1 Vito Volpetti,2 George P. Lomonossoff,2 Kurt Nielsen3 and Laurence D. Barron1

1 Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK
2 Department of Metabolic Biology, John Innes Centre, Norwich NR4 7UH, UK
3 Department of Chemistry, DTU 207, Technical University of Denmark, DK-2800 Lyngby, Denmark

A vibrational Raman optical activity (ROA) study of a range of different structural types of virus exemplified by filamentous bacteriophage fd, tobacco mosaic virus, satellite tobacco mosaic virus, bacteriophage MS2 and cowpea mosaic virus has revealed that, on account of its sensitivity to chirality, ROA is an incisive probe of their aqueous solution structures at the molecular level. Protein ROA bands are especially prominent from which, as we have shown by comparison with the ROA spectra of proteins with known structures and by using a pattern recognition program, the folds of the major coat protein subunits may be deduced. Information about amino acid side-chain conformations, exemplified here by the determination of the sign and magnitude of the torsion angle $\chi^2$ for tryptophan in fd, may also sometimes be obtained. By subtracting the ROA spectrum of the empty protein capsid (top component) of cowpea mosaic virus from those of the intact middle and bottom-upper components separated by means of a caesium chloride density gradient, the ROA spectrum of the viral RNA was obtained, which revealed that the RNA takes up an A-type single-stranded helical conformation and that the RNA conformations in the middle and bottom-upper components are very similar. This information is not available from the X-ray crystal structure of cowpea mosaic virus since no nucleic acid is visible.

Introduction

Knowledge of the structures of viruses at the molecular level is essential for understanding their modus operandi and hence for enterprises such as structure-guided antiviral drug design (Chiu et al., 1997). However, the widespread application of key structural biology techniques to intact viruses is often hampered by practical difficulties. Although of immense value, X-ray structures at near atomic resolution are only available for several dozen of the thousands of different viruses currently recognized; and to study the molecular details of larger and more complex viruses, it is necessary to ‘dissect’ them into well-defined subunits or substructures (Harrison, 2001). Even then, only the coat protein subunit structures are usually fully resolved, much of the nucleic acid often being too disordered to provide useful diffraction data.

Conventional Raman spectroscopy is already valuable in structural virology owing to its ability to provide information about protein and nucleic acid constituents of intact viruses by means of their characteristic vibrational bands (Thomas, 1987, 1999). On account of its sensitivity to molecular chirality a novel form of Raman spectroscopy called Raman optical activity (ROA), one version of which measures vibrational optical activity by means of small differences in the Raman spectra of chiral molecules using right- and left-circularly polarized incident laser light (Nafie, 1997; Barron & Hecht, 2000), has recently been developed into a new probe of biomolecular structure (Barron et al., 2000). ROA is more incisive than conventional vibrational spectroscopy in the study of biomolecules such as proteins, nucleic acids and carbohydrates because the few vibrational coordinates which sample the skeletal chirality most directly, such as those within the peptide backbone of a protein, make the largest contributions, yielding ROA band patterns which are simpler than the parent Raman band patterns and more sensitive to conformation (Barron et al., 2000). In proteins the many discrete ROA bands, originating in loops and turns as well as secondary structure, enable fold information to be deduced.

Early studies on filamentous bacteriophages (Blanch et al., 1999, 2001a) and helical plant viruses (Blanch et al., 2001b, 2002) demonstrated that intact viruses are accessible to ROA
measurements from which new information about coat protein structures may be obtained. The work presented here provides a comparative ROA study of filamentous bacteriophage fd, tobacco mosaic virus (TMV), satellite tobacco mosaic virus (STMV), the empty protein capsid of bacteriophage MS2 and cowpea mosaic virus (CPMV) in order to validate the ability of ROA to adequately discriminate between different structural types and to obtain new molecular information in some cases. These particular samples were chosen because they cover a range of structures (filamentous, rigid rod and icosahedral, the latter not having been previously studied with ROA), some information is already available about the structures at both particle and molecular level, and sufficient quantities of material were available. In addition, the ROA study of CPMV provided an opportunity to deduce information on its nucleic acid structure about which nothing previously was known since no nucleic acid is visible in the refined X-ray crystal structure (Lin et al., 1999).

**Methods**

Bacteriophage fd was prepared using standard methods (Smith & Scott, 1993). STMV was purified from leaves of infected tobacco plants (Nicotiana tabacum) (Koszelak et al., 1989). The empty MS2 protein capsids were prepared as described (Stonehouse & Stockley, 1993). The preparation of CPMV and its separation into top, middle and bottom components by caesium chloride density gradient centrifugation at pH 7–0 followed standard procedures (Lin et al., 1999). Rabbit calcyclin (S100 A6) was prepared as described (Pedrocchi et al., 1994). Jack bean concanavalin A was supplied by Fluka.

The virus solutions, at \( \sim 10–20 \) mg/ml, were held in small rectangular quartz microfluorescence cells at ambient temperature (\( \sim 20^\circ\mathrm{C} \)) except for STMV, which was studied at 4 \( ^\circ\mathrm{C} \) to increase its stability. The peptide and protein concentrations were \( \sim 50–70 \) mg/ml. The ROA measurements were performed using an instrument described previously (Hecht et al., 1999). 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, 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 laser light, respectively. The experimental conditions were: laser wavelength 514.5 nm; laser power at the sample \( \sim 700 \) mW; spectral resolution \( \sim 10 \) cm\(^{-1} \); acquisition times \( \sim 72 \) h for the viruses and \( \sim 10 \) h for the peptide and proteins. A shallow focus of the laser beam through the samples was used to avoid sample degradation, which was negligible over the duration of the measurements as judged by the absence of any changes in the parent Raman spectra.

**Results and Discussion**

**Filamentous bacteriophage fd**

The fd particle takes the form of a flexible rod \( \sim 6 \) nm wide and \( \sim 880 \) nm long comprising a single-stranded DNA genome containing \( \sim 6500 \) nucleotides (\( \sim 12\% \) of the particle mass) surrounded by a cylindrical shell consisting of a helical array of \( \sim 2700 \) identical major coat proteins each containing 50 amino acids (Marvin, 1998). X-ray fibre diffraction reveals that the major coat proteins have the fold of an extended \( \alpha \)-helix and that they overlap like fish scales with \( \sim \) half of each protein subunit exposed on the surface and the other half protected (Marvin, 1998).
helix ~ helix interfaces of the overlapping major coat proteins and to the exposed hydrated surface residues, respectively (Blanch et al., 1999; Barron et al., 2000).

The positive ROA band of fd at ~ 1559 cm⁻¹ originates in a W3-type vibration of the single tryptophan side-chain in the major coat proteins. The wavenumber of the W3 band in the conventional Raman spectra of proteins has been used to deduce the magnitude of the torsion angle $\chi^{2,1}$, which was found to be [120 °] for fd (Miura & Thomas, 1995). However, the sign of the W3 ROA band gives in addition the sign of $\chi^{2,1}$ (Blanch et al., 2001a). Since the ~ 1559 cm⁻¹ W3 ROA band in fd is positive, we deduce that $\chi^{2,1} \sim + 120 °$. This angle is not usually available from the X-ray fibre diffraction data. Experimental determination of the sign as well as the magnitude of $\chi^{2,1}$ will aid its application to studies of tryptophan environments in filamentous bacteriophages (Miura & Thomas, 1995) and to the development of molecular models (Tsuboi et al., 1996).

### Tobacco mosaic virus

The U1 strain of TMV studied here comprises single-stranded positive-sense RNA containing ~ 6400 nucleotides (~ 5% of the particle mass) within a rigid helical rod-shaped particle ~ 18 nm wide and ~ 300 nm long with ~ 2130 identical coat proteins each containing 158 amino acids. X-ray fibre diffraction reveals the coat protein subunits to be based on a four-helix bundle (Namba et al., 1989).

The Raman and ROA spectra of the U1 strain of TMV, taken from an earlier study (Blanch et al., 2002), are presented in Fig. 1(b). With positive extended amide III bands at ~ 1295 and 1342 cm⁻¹ assigned to hydrophobic and hydrated $\alpha$-helix, respectively, together with the characteristic amide I couplet plus other features, the ROA spectrum of TMV has the general appearance expected for helix bundle proteins such as that of rabbit calcyclin (S100 A6), presented in Fig. 2(b). ROA bands from side-chains can be identified, including the positive band at ~ 1554 cm⁻¹ assigned to the three tryptophans.

### Satellite tobacco mosaic virus

STMV is a small T = 1 icosahedral virus of diameter ~ 17 nm comprising single-stranded RNA containing 1059 nucleotides (~ 25% of the particle mass) inside a capsid composed of 60 identical copies of a coat protein with 159 amino acids. The X-ray crystal structure of STMV ( Larson et al., 1998) reveals the coat protein fold to be that of a jelly roll $\beta$-barrel (sandwich). Approximately 45% of the RNA is visible and takes the form of stem–loop elements (Larson & McPherson, 2001).

The Raman and ROA spectra of STMV are presented in Fig. 1(c). The ROA spectrum contains prominent bands characteristic of $\beta$-sheet (Barron et al., 2000), such as the negative-positive amide I couplet centred at ~ 1665 cm⁻¹ and the strong sharp negative ROA band at ~ 1246 cm⁻¹. The
negative band at \( \sim 1369 \text{ cm}^{-1} \) is characteristic of hairpin bends (Barron et al., 2000). The strong positive ROA band at \( \sim 1316 \text{ cm}^{-1} \) may originate in polyproline II (PPII)-helical elements in some of the loops and the long N-terminal disordered strand (Barron et al., 2000). The overall appearance of the STMV ROA spectrum is similar to those of proteins with the jelly roll fold such as jack bean concanavalin A, shown in Fig. 2(c). The weak positive ROA bands at \( \sim 1483 \) and \( 1574 \text{ cm}^{-1} \) may arise from the RNA bases guanine and adenine (Thomas, 1999; Bell et al., 1998). Some of the protein ROA bands may be distorted by contributions from RNA bands, especially the negative band at \( \sim 1246 \text{ cm}^{-1} \) since RNAs show strong negative ROA near this wavenumber (vide infra).

**Bacteriophage MS2**

MS2 is a small \( T = 3 \) icosahedral virus of diameter \( \sim 25 \text{ nm} \) comprising single-stranded RNA containing 3569 nucleotides (\( \sim 25\% \) of the particle mass) inside a capsid composed of 180 identical copies of a coat protein containing 129 amino acids. The X-ray crystal structure reveals that the coat protein subunits fold into a five-stranded antiparallel \( \beta \)-sheet with an additional short hairpin at the N terminus and two \( \alpha \)-helices at the other end of the chain (Valegård et al., 1990). The \( \alpha \)-helices are responsible for interactions with a second subunit to form a dimer containing 10 adjacent antiparallel \( \beta \)-strands. This fold is quite different from the jelly roll found in most other families of icosahedral viruses to date.

The Raman and ROA spectra of the empty protein capsid of MS2 are presented in Fig. 1(d). Although the ROA spectrum is much more similar to that of STMV than to those of fd and TMV, there are some differences of detail which reflect the two distinct types of \( \beta \)-sheet fold. Unfortunately, a model protein with a fold similar to that of the coat protein subunit of MS2, such as the peptide-binding domain of a class I MHC protein (Branden & Tooze, 1999), was unavailable to us.

**Cowpea mosaic virus: separation of protein and nucleic acid ROA spectra**

CPMV is the type member of the comovirus group of plant viruses. It has a genome consisting of two molecules of positive-sense RNA (RNA-1 and RNA-2) which are separately encapsidated in \( T = 3 \) icosahedral particles of diameter \( \sim 28 \text{ nm} \), the structure of which is known to atomic resolution (Lomonossoff & Johnson, 1991; Lin et al., 1999, 2000). Virus preparations can be separated into four components, designated top (T), middle (M), bottom-upper (\( B_U \)) and bottom-lower (\( B_L \)) by centrifugation on density gradients. The four components have identical protein compositions and contain 60 copies each of a large (L) and small (S) coat protein made up of 374 and 213 amino acids, respectively. The T-component is devoid of RNA, the M-component contains a single molecule of RNA-2 made up of 3481 nucleotides (\( \sim 24\% \) of the particle mass), and the \( B_U \)-component contains a single molecule of RNA-1 made up of 5889 nucleotides (\( \sim 34\% \) of the particle mass). The \( B_L \)-component contains a single molecule of RNA-1, plus a number of \( \text{Cs}^+ \) ions which have permeated the virion from the CsCl gradient used to separate the components. The X-ray crystal structure of CPMV reveals the folds of the L and S proteins to be of the jelly roll \( \beta \)-barrel (sandwich) type (Lomonossoff & Johnson, 1991; Lin et al., 1999), but no nucleic acid is visible (Lin et al., 1999).

The Raman and ROA spectra of the T-, M- and \( B_U \)-components of CPMV are presented in Fig. 3(a, b, c) respectively. The overall appearance of the ROA spectrum of T-CPMV is very similar to those of proteins with the jelly roll fold, such as jack bean concanavalin A (Fig. 2c). However, the ROA spectra of M- and \( B_U \)-CPMV contain a number of extra bands, some of which are quite intense. Subtraction of the ROA spectrum of T-CPMV from those of M-CPMV and \( B_U \)-
Molecular structures of viruses

CPMV produces the spectra presented in Fig. 4(a, b), respectively. These difference spectra are very similar to those of synthetic and natural RNAs published previously (Bell et al., 1997, 1998) and we therefore take them as originating mainly in the viral RNA. The ROA spectrum of magnesium-free phenylalanine-specific transfer RNA (tRNAPhe), taken from Bell et al. (1998), which adopts an open clover leaf (secondary structure) form, is reproduced in Fig. 4(c) for convenience. The clear negative-positive-negative triplet with peaks at ~ 996, 1050 and 1094 cm⁻¹ in the two difference spectra is very similar to that in tRNAPhe where it was assigned to sugar-phosphate vibrations associated with the C3'-endo sugar ring pucker as found in A-type helices (Bell et al., 1997, 1998). The ROA couplet, negative at ~ 1669 and positive at ~ 1702 cm⁻¹ in the two difference spectra, may have a similar origin to the weaker couplet in the same region of the ROA spectrum of tRNAPhe which was assigned to ring vibrations of bases involved in the base stacking interactions found in A-type helices (Bell et al., 1998). The ROA band pattern in the region ~ 1200–1410 cm⁻¹ in the two difference spectra is very similar to that in tRNAPhe in the same region: this band pattern was assigned to vibrations involving mixing of vibrational coordinates from both base and sugar rings which reflect the mutual orientation of the sugar and base rings found in A-type helices (Bell et al., 1998).

The (M – T)- and (B_U – T)-CPMV difference ROA spectra displayed in Fig. 4(a, b) therefore contain distinct features previously associated with A-type helices, so we deduce that much of the viral RNA in both M- and B_U-CPMV adopts an A-type helical conformation. The absence of similar bands to those observed at ~ 1480 and 1530 cm⁻¹ in tRNAPhe may reflect a single-strand helix since these bands possibly originate in vibrations of bases within the base-paired regions of tRNAPhe. The only anomalous feature is the strong broad ROA band peaking at ~ 1455 cm⁻¹, the origin of which is not clear at present. The close similarity of the two difference spectra reveals that the RNA adopts very similar conformations in the M- and B_U-components: any differences are too small to be reliable at the present level of sensitivity. Unlike the X-ray crystal structure of CPMV, or of any other comovirus (Lin et al., 2000), that of the middle but not the bottom component of bean pod mottle virus (BPMV) does reveal some RNA structure which is found to have stacked bases and a backbone conformation approximately like that found for one strand of an A-type RNA helix but no base pairing (Chen et al., 1989; Lomonossoff & Johnson, 1991). Our results therefore suggest that the RNA conformations in the middle and bottom-upper components of CPMV may be similar to that observed in the middle component of BPMV.

Principal component analysis

We are developing a pattern recognition program, based on principal component analysis (PCA), to identify protein folds from ROA spectral band patterns (Blanch et al., 2002). From the ROA spectral data, PCA calculates a set of subspectra that serve as basis functions which can be used to reconstruct any member of the original set of ROA spectra. Fig. 5 shows a plot of the coefficients for our current set of 73 polypeptide, protein and virus ROA spectra for the two most important basis functions, together with molscript diagrams (Kraulis, 1991) of the four basic virus coat protein subunit folds being considered. The protein positions 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 quantitative estimates of structural elements such as helix, sheet, loops and turns. The spectra separate into clusters corresponding to different types of protein structure, with increasing α-helix content to the left,
Fig. 5. Plot of the PCA coefficients for the two most important basis functions for a set of 73 polypeptide, protein and virus ROA spectra. The positions of the ROA spectra under discussion fall in regions consistent with their known structures. Models of the four basic virus coat protein subunit folds were drawn from PDB structures 1ifj (fd), 1vtm (TMV), 1a34 (STMV) and 1msc (MS2). More complete definitions of the structural types are as follows: all alpha, >~ 60% α-helix with little or no other secondary structure; mainly alpha, >~ 35% α-helix and a small amount of β-sheet (~5–15%); alpha beta, similar significant amounts of α-helix and β-sheet; mainly beta, >~ 35% β-sheet and a small amount of α-helix (~5–15%); all beta, >~ 45% β-sheet with little or no other secondary structure; mainly disordered/irregular, little secondary structure; all disordered/irregular, no secondary structure.

Increasing β-sheet content to the right, and increasing disordered or irregular structure from bottom to top. Even without quantitative estimates of structural elements, since ROA spectra contain many more structure-sensitive bands than other spectroscopies such as ultraviolet circular dichroism and Fourier-transform infrared, PCA analysis of ROA spectra invariably provides better discrimination between different structural types of protein.

The presence of nucleic acid bands in the ROA spectra of intact viruses will affect their positions in the plot, but 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-, M- and B1−CPMV all fall closely together in the plot despite the ROA spectra of M- and B1−CPMV containing a number of strong RNA bands that are absent from that of T-CPMV.

Bacteriophage fd lies well inside the all alpha region, a little distance from the bottom and level with α-helical poly(L-lysine), reflecting the long stretches of α-helix but with some irregular structure. TMV lies within the left side of the mainly alpha region and some distance from the bottom, consistent with the helix bundle fold plus a little β-sheet with some long loops. Calcyclin falls between fd and TMV. STMV is located within the mainly beta region and some distance from the bottom, consistent with its jelly roll fold containing mainly β-sheet plus a small amount of 310-helix and some long loops. The MS2 capsid also lies within the mainly beta region, but closer to the bottom, consistent with its unusual fold containing long stretches of flat antiparallel β-sheet plus some α-helix, but fewer loops. T-, M- and B1−CPMV fall on the left side of the all beta region, consistent with their jelly roll fold containing a large amount of β-sheet plus a small amount of α-helix, and some way from the bottom consistent with the presence of a number of loops. Concanavalin A appears at the right side of the beta region and some distance from the bottom, consistent with a large amount of β-sheet, no α-helix and many loops.

Conclusions
This study has shown that, despite the enormous size and complexity of viruses, the special sensitivity of ROA to skeletal chirality results in simple ROA spectra that readily differentiate between different virus types and which provide
information about the structures of the protein and nucleic acid components. New information was obtained about the tryptophan side-chain conformation in the major coat proteins of fd and about the RNA conformation in the middle and bottom-upper components of CPMV.

Pattern recognition methods together with inspection of ROA spectral details should, among other things, provide the major coat protein subunit folds of most viruses of unknown structure. There may be sufficient sensitivity to characterize small conformational changes of the capsid subunits as the supramolecular assembly undergoes structural maturation. When the empty protein capsid is available, subtraction of its ROA spectrum from that of the intact virus provides the ROA spectrum of the nucleic acid (perhaps with small extra features reflecting protein conformational changes due to protein–nucleic acid interactions). However, 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. PCA may also yield the ROA spectrum of carbohydrate moieties. The ability of ROA to provide information about the nucleic acid structure within an intact virus particle augments a similar ability of conventional Raman spectroscopy (Thomas, 1987, 1999) and should be especially valuable since X-ray diffraction has been rather uninformative in this area with few examples of clearly resolved RNA or DNA structures. Our results suggest that ROA should find many applications in structural virology.

We thank the following for their generous gifts of virus or protein samples: S. J. Holland and R. N. Perham, Department of Biochemistry, University of Cambridge (fd); A. McPherson, Department of Molecular Biology and Biochemistry, University of California, Irvine (STMV); N. J. Stonehouse and P. G. Stockley, Astbury Centre for Structural Molecular Biology, University of Leeds (MS2 capsid); and G. Fritz, Institute of Biochemistry, University of Zürich (calcyclin). We are grateful to the BBSRC for a research grant, the EPSRC for a Research Studentship for C.D.S., and the John Innes Foundation for a Studentship for V.V.

References


Received 17 May 2002; Accepted 10 June 2002