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

Ewan W. Blanch,1 David J. Robinson,2 Lutz Hecht,1 Christopher D. Syme,1 Kurt Nielsen3 and Laurence D. Barron1

1 Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK
2 Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK
3 Department of Chemistry, DTU 207, Technical University of Denmark, DK-2800 Lyngby, Denmark

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

Introduction

Potato virus X (PVX) and narcissus mosaic virus (NMV) belong to the *Potexvirus* genus of helical plant viruses, PVX being the type species (Brunt et al., 2000). They contain genomes of single-stranded positive-sense RNA encapsidated within flexuous filamentous particles, \(~500\) nm long and \(~13–14\) nm in diameter, with helical symmetry and a pitch of \(~3.4\) nm. In comparison, the *Tobamovirus* genus of helical plant viruses, to which tobacco mosaic virus (TMV) belongs, also contain single-stranded positive-sense RNA but the particles are rigid rods \(~300\) nm long and \(~18\) nm in diameter with a helical pitch of \(~2.4\) nm. Although detailed information about the genome organization, the amino acid sequences of the encoded proteins and the sequence homology between members of the *Potexvirus* genus is available (Brunt et al., 2000) and references therein, little is known about the conformation of the coat protein subunits and their packing arrangements in the virus particles. Unlike the *Tobamoviruses*, there are no X-ray diffraction data of sufficient resolution to show molecular conformational detail, but ultraviolet circular dichroism (UVCD) measurements suggest \(~45\%\) \(\alpha\)-helix content and \(~5\%\) \(\beta\)-sheet for both PVX (Homer & Goodman, 1975) and NMV (Wilson et al., 1991). A model of the PVX coat protein tertiary structure has been suggested based on the predicted secondary structure content, the principles of secondary structure packing, and the known biochemical, immunological and tritium bombardment data (Baratova et al., 1992a).

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

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

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

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

Methods

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

Both viruses were studied as solutions at $\sim 25 \text{mg/ml}$ in 15 mM phosphate buffer at pH $7.4$ held in small quartz microfluorescence cells at ambient temperature ($\sim 20 ^\circ \text{C}$). The ROA measurements were performed using an instrument described previously (Hecht et al., 1999). It has a backscattering configuration and employs a single-grating spectrograph fitted with a backthinned charge coupled device (CCD) camera as detector and an edge filter to block the Rayleigh line. The small ROA signals are accumulated by synchronizing the Raman spectral acquisition with an electrooptic modulator that switches the polarization of the argon-ion laser beam between right- and left-circular at a suitable rate. The spectra are displayed in analogue-to-digital counter units as a function of the Stokes Raman wavenumber shift with respect to the exciting laser wavelength. The ROA spectra, which have undergone minimal smoothing (2 point FFT), are presented as circular intensity differences $\Delta R^{+} - \Delta R^{-}$ and the parent Raman spectra as circular intensity sums $\Delta R^{+} + \Delta R^{-}$, where $\Delta R^{+}$ and $\Delta R^{-}$ are the Raman-scattered intensities in right- and left-circularly polarized incident light, respectively. The experimental conditions were as follows: laser wavelength 514.5 nm; laser power at the sample $\sim 700 \text{mW}$; spectral resolution $\sim 10 \text{cm}^{-1}$; acquisition time $\sim 48 \text{h}$.

Results and Discussion

The ROA spectral details and their structural assignments

The backscattered Raman and ROA spectra of PVX and NMV are shown as the top and middle pairs of spectra,
Fig. 1. Backscattered Raman ($I^+ - I^-$) and ROA ($I^+ + I^-$) spectra of PVX (top pair), NMV (middle pair) and TMV (bottom pair) in phosphate buffer at pH 7.4.

respectively, in Fig. 1. An account of most of the ROA assignments given in what follows can be found in Barron et al. (2000). The ROA spectra of the two viruses are dominated by protein bands. The general appearance of the ROA spectra of both viruses suggests that both contain a large amount of \(\alpha\)-helix. In particular, both show an ROA couplet, negative at ~1630 and positive at ~1666 cm\(^{-1}\), originating in peptide amide I vibrations involving mainly C=O stretching that is characteristic of \(\alpha\)-helix. Positive protein ROA bands in the range ~1295–1312 cm\(^{-1}\) together with positive bands in the range ~1337–1345 cm\(^{-1}\), originating in extended amide III vibrations involving various contributions from peptide backbone N–H and C\(_\alpha\)–H deformations and the C\(_\alpha\)–N stretch, are also characteristic of \(\alpha\)-helix. The former appear to arise from \(\alpha\)-helix in a hydrophobic environment and the latter to a hydrated form of \(\alpha\)-helix (Blanch et al., 1999). Both PVX and NMV show strong positive ROA bands at ~1337 cm\(^{-1}\) together with weaker positive intensity centred ~1295–1300 cm\(^{-1}\) which are therefore assigned to hydrated and hydrophobic \(\alpha\)-helix, respectively. Such an ROA band pattern is characteristic of helix bundle proteins in which most residues are surface exposed and hydrated and a smaller number are within hydrophobic interfaces between helices. One difference between the ROA spectra of the two viruses is the presence of a significant negative band at ~1252 cm\(^{-1}\) in the ROA spectrum of NMV assigned to \(\beta\)-strand, whereas PVX shows rather weaker less well-defined negative ROA here. However, both viruses show negative ROA bands at ~1220 cm\(^{-1}\) assigned to a hydrated form of \(\beta\)-strand. Differences in various weak bands reflect differences in the details of loop and turn sequences.

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

The protein ROA data are consistent with the UVCD results, which predicted large amounts of \(\alpha\)-helix with little \(\beta\)-sheet (Homer & Goodman, 1975; Wilson et al., 1991). Furthermore, the overall appearances of the ROA spectra of PVX and NMV are very similar to each other and to that of TMV reported recently (Blanch et al., 2001a) for which the coat protein subunits are based on four-helix bundles containing both water-exposed residues and residues at hydrophobic helix–helix interfaces (Namba et al., 1989; Stubbs, 1999). One difference, however, is that the positive ~1337 cm\(^{-1}\) ROA bands in PVX and NMV are significantly more intense than the corresponding band in TMV, which suggests that PVX and NMV contain more hydrated \(\alpha\)-helix than TMV. For convenience, a re-measured ROA spectrum of TMV (strain U1) of better quality than that recorded in the earlier study is shown at the bottom of Fig. 1. Our results therefore suggest that the coat protein folds of PVX and NMV are similar to that of TMV, being based on a helix bundle, but with differences of detail resulting from differences in the appended loops and turns and from the extra sequences present in PVX and NMV (the coat proteins of the DX strain of PVX, the MC strain of
NMV and the U1 strain of TMV contain 237, 240 and 159 amino acids, respectively). Our results do not support the model of Baratova et al. (1992a) for the PVX coat proteins, which may contain too much well-defined β-sheet (which is predicted at both the N- and C-terminal ends). However, unlike the N-terminal regions of PVX which tritium planigraphy shows are exposed, the C-terminal regions of PVX are virtually inaccessible to tritium labelling in the intact virion (Baratova et al., 1992b), but become accessible after ~20 N-terminal amino acids are proteolytically removed (Baratova et al., 1992a). For TMV, on the other hand, tritium planigraphy shows that both N- and C-terminal regions are exposed (Goldanski et al., 1988), as expected from the X-ray structure (Stubbs, 1999). Some of the additional ~80 residues in PVX and NMV compared with TMV might constitute an N-terminal region which tends to bury the C-terminal end, but the ROA spectra suggest that this region is not organized into a β-sheet as proposed by Baratova et al. (1992a). In this connection it is worth noting that there is no similarity in primary sequence of PVX and TMV coat proteins since no alignment of the potexvirus and tomatovirus conserved primary sequences is possible and also that, according to a prediction of Sawyer et al. (1987), the N-terminal 50 residues of PVX have zero probability of being α-helix. This last prediction is unfortunate since it is tempting to assign the greater hydrated α-helix content of PVX and NMV relative to TMV to the extra sequences.

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

A few ROA bands from side-chains can also be identified. ROA bands in the range ~1420−1480 cm⁻¹ originate in both aliphatic and aromatic side-chains. Proteins often show a tryptophan ROA band in the ~1550 cm⁻¹ region, assigned to a W3 type vibration of the indole ring, which reflects the sign and magnitude of the torsion angle χ2,1 (Blanch et al., 2001b). The PVX and NMV strains used here contain five and four tryptophans, respectively. The absence of significant ROA in this spectral region of both viruses therefore suggests conformational heterogeneity among these tryptophans (Blanch et al., 2001b). On the other hand, the two and six tyrosines present in PVX and NMV, respectively, could be responsible for the small but significant negative ROA bands at ~1604 cm⁻¹ associated with Y8a,b type modes of the aromatic ring (Miura & Thomas, 1995), suggesting these side-chains have mainly fixed conformations generating ROA signals which tend to reinforce. A similar negative band is present at ~1608 cm⁻¹ in the ROA spectrum of TMV, which contains four tyrosines. These negative Y8 ROA bands may originate in the participation of some tyrosine side-chains in subunit binding, as observed in the TMV X-ray fibre structure of Namba et al. (1989).

Principal component analysis of coat protein folds

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

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

The overall structures of PVX and NMV

Despite the lack of significant amino acid sequence homology to the TMV coat protein, Sawyer et al. (1987) suggested on the basis of predicted α-helical segments that the PVX coat protein exhibits an approximate twofold symmetry between the amino and carboxyl halves of the molecule such that the α-helices align themselves in pairs like in the TMV coat protein. This reinforces our conclusion from the ROA data that the PVX, NMV and TMV coat protein folds are similar. Furthermore, the low resolution X-ray fibre diffraction data available for several potexviruses are consistent with α-helices packed approximately at right angles to the viral axis, as in TMV (Stubbs, 1989). Hence the helical architecture of the PVX and NMV virus particles may be similar to that of TMV, but with different structural parameters. These different structural parameters may be associated with the small differences in the coat protein folds of PVX and NMV compared with TMV indicated by this study and which, together with their extra sequences, would lead to differences in packing of the protein subunits.

We thank the BBSRC for a research grant and the EPSRC for a Research Studentship for C.D.S. The Scottish Crop Research Institute is grant-aided by the Scottish Executive Environment and Rural Affairs Department.

References


Committee on Taxonomy of Viruses

The TMV protein-accessible surface of the virus. 


