Shelnutt, J. A., Rousseau, D. L., Dethmers, J. K., & Margoliash, E. (1979a) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3865.
Shelnutt, J. A., Rousseau, D. L., Friedman, J. M., & Simon, S. R. (1979b) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4409.
Shire, S. J. (1974) Ph.D. Thesis, Indiana University, Bloomington, IN.
Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1974a) *Biochemistry 13*, 2967.
Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1974b) *Biochemistry 13*, 2974.
Sober, H. A., & Harte, R. A., Eds. (1968) *Handbook of Biochemistry, Selected Data for Molecular Biology*, p J-51, Chemical Rubber Co., Cleveland, OH.
Sturtevant, J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236.
Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Chapter 8, Wiley, New York.
Tanford, C. (1980) *The Hydrophobic Effect*, 2nd ed., Wiley, New York.
Tanford, C., & Kirkwood, J. (1957) *J. Am. Chem. Soc. 79*, 5333.
Thomas, J. O., & Edelstein, S. J. (1973) *J. Biol. Chem. 248*, 2901.
Valdes, R., & Ackers, G. K. (1977) *J. Biol. Chem. 252*, 74.
Van Beek, G. G. M., & deBruin, S. H. (1980) *Eur. J. Biochem. 105*, 353.
Van Beek, G. G. M., Zuiderweg, E. R. P., & deBruin, S. H. (1979) *Eur. J. Biochem. 99*, 379.
Viggiano, G., & Ho, C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3673.
Woodward, C. K., & Hilton, B. D. (1979) *Annu. Rev. Biophys. Bioeng. 8*, 99.

Structural Studies of Tobacco Mosaic Virus and Its Components by Laser Raman Spectroscopy†

S. R. Fish, K. A. Hartman,* G. J. Stubbs, and G. J. Thomas, Jr.*

ABSTRACT: Raman spectra have been recorded for oriented gels of tobacco mosaic virus, aqueous solutions of TMV protein in the form of capsid and disk, and protein-free TMV RNA. The protein-free RNA molecule contains about 80% ordered type-A structure, which is similar to other RNA molecules in solution. RNA in the virus particle does not exhibit the base stacking and pairing normally found in protein-free RNA but instead shows at least two distinct structures characterized by different phosphodiester group frequencies in the Raman spectrum. Approximately one out of three such linkages retains the type-A geometry characteristic of aqueous RNA, while two out of three appear to have a different geometry not encountered previously for RNA. The secondary structure of the TMV coat protein molecule is similar for the three aggregation states studied (virus, capsid, and disk) and consists of 40–50% α helix, 40–50% irregular structure, and 0–20% β sheet. Two of the three tryptophan residues per protein molecule reside in hydrophobic regions, and the third is in contact with water. The distribution of the four tyrosine residues per protein molecule among hydrogen-bonding states is the same for capsid and virus but different for the disk. The latter could be a consequence of the presence of Cl^- ions which may form hydrogen bonds with one or more *p*-hydroxyl groups of tyrosine residues.

The tobacco mosaic virus (TMV) particle, or virion, is a relatively simple system in which many copies of one kind of protein molecule are combined in a helical array with one strand of RNA to form a rod-shaped structure. TMV protein, which can be separated from the RNA, forms many different aggregates including a disk which is composed of 34 protein molecules and a rod-shaped capsid which has the same helical arrangement of protein subunits found in the virion. The disk and protein-free RNA interact to form the virion in vitro and also, it is believed, in vivo. The RNA, disk, capsid, and virion compose a system in which details of protein–protein and protein–RNA interactions may be investigated (Caspar, 1963; Holmes, 1980).

The TMV system has received considerable study by numerous techniques of biochemistry and biophysics, and recent reviews are available (Kaper, 1975; Holmes, 1980). In particular, X-ray diffraction studies have produced electron density maps at 4-Å resolution for the virion (Stubbs et al., 1977) and at 2.8-Å resolution for the disk (Bloomer et al., 1978). Molecular models derived from these maps show many details of the molecular structures and interactions.

Raman spectroscopy measures the scattering of light as modulated by the normal modes of vibration in molecules of the sample. The intensities and frequencies of the Raman lines can be related to the interactions and structures of the molecular subgroups involved and in favorable cases have been used to measure protein and nucleic acid secondary structures. These results have permitted the use of Raman spectroscopy for the study of structure of both RNA and DNA viruses and nucleoproteins (Thomas, 1978).

It was the goal of this work to obtain Raman spectra of TMV and its disk, capsid, and RNA and to interpret these in terms of protein and RNA structures and interactions. Two papers have recently appeared in which Raman spectra of TMV (Shie et al., 1978) and additionally some of its com-

† From the Department of Biochemistry and Biophysics, University of Rhode Island, Kingston, Rhode Island 02881 (S.R.F. and K.A.H.), Rosenstiel Basic Sciences Center, Brandeis University, Waltham, Massachusetts 02254 (G.J.S.), and Department of Chemistry, Southeastern Massachusetts University, North Dartmouth, Massachusetts 02747 (G.J.T.). *Received July 22, 1981.* Paper IX in the series Studies of Virus Structure by Laser Raman Spectroscopy. For paper VIII in this series, see Thomas & Day (1981). Supported in part by U.S. Public Health Service Grant AI 11855 (G.J.T.) and by the Department of Biophysics, University of Rhode Island (K.A.H.).

0006-2960/81/0420-7449$01.25/0

ponents (Fox et al., 1979) were presented. The previous Raman data on TMV will be discussed later in this paper, in connection with the present results.

Materials and Methods

Samples of TMV were kindly supplied as stock solutions of the *Vulgare* strain by Dr. C. A. Knight (University of California, Berkeley) and Dr. K. A. Holmes (Max-Planck-Institute, Heidelberg).

TMV protein (A protein) was isolated from the virus by the method of Fraenkel-Conrat (1957). The capsid was prepared from solutions of A protein in distilled H_2O by increasing the ionic strength to 0.05 M with 2 M KCl and lowering the pH to 5.5 with 0.1 M HCl. After incubation for 16 h at 4 °C, the capsid was harvested by centrifugation at 189000*g* for 100 min by using a Beckman SW50.1 rotor. The resulting pellet was dissolved in distilled water to a concentration of about 60 mg/mL.

Disks were prepared from solutions of A protein by increasing the ionic strength to 0.8 with KCl. After 16 h, the solution was centrifuged for 100 min at 189000*g* to remove large aggregates of protein, after which the disks remaining in the supernatant were pelleted by centrifugation at 189000*g* for 280 min. The pellet was dissolved in 0.8 M KCl, pH 8, to a concentration of about 50 mg/mL.

TMV RNA was purified by the phenol extraction method of Knight (1975) which prescribes the use of 0.02 M phosphate buffer during the deproteination steps. However, when ethanol is subsequently added to precipitate the RNA, phosphate salts are also precipitated. For removal of the phosphate salts, which cause intense interfering lines in the Raman spectrum, the pellet was redissolved in H_2O and exhaustively dialyzed against distilled H_2O. The resulting RNA solution was made 0.1 M with NaCl and precipitated with 2 volumes of ethanol. The resulting pellet was washed twice with cold 70% ethanol–water, dried with a stream of air, and dissolved in 1 mM NaCl to approximately 30 mg/mL. (In an alternative procedure, better suited to Raman spectroscopy, phosphate buffer may be replaced with 0.1 M NaCl or KCl which allows the precipitation of RNA without coprecipitation of interfering phosphate salts.)

The total protein concentration in capsid, disk, or A protein was measured spectrophotometrically by using the extinction coefficient at 280 nm, ϵ_{280} = 1.27 mL mg^{-1} cm^{-1} (Fraenkel-Conrat, 1957). All protein samples exhibited a UV absorption band maximum at 280 nm and minimum at 260 nm, as expected (Durham, 1972). Concentrations of TMV and RNA were measured by using ϵ_{260} values of 2.7 and 24 mL mg^{-1} cm^{-1}, respectively.

The integrity of TMV samples and the success of the preparations of capsids and disks were evaluated by using UV spectroscopy and sedimentation. Samples of TMV were layered on 5–20% sucrose gradients and sedimented by using the Beckman SW50.1 rotor. A reasonably narrow band was obtained between 150 and 200 S as measured relative to a 70S marker band from *Escherichia coli* ribosomes.

For the capsid, sucrose-gradient centrifugation showed a broad region of sedimentation above 50 S as measured from *E. coli* ribosomal subunits. This is expected since capsids prepared as above have varying lengths (Durham, 1972; Durham & Finch, 1972).

The capsid is clearly distinguished from the A protein by its characteristic light scattering which can be observed in the near-UV spectrum (Bakke, 1979). Our samples of capsid show this expected light scattering between 350 and 310 nm. Samples of A protein, on the other hand, show little or no such scattering in the 350–310-nm interval.

Raman spectra were measured with a Spex Ramalog spectrometer using for excitation the line at either 488.0 or 514.5 nm from an argon ion laser (Coherent Radiation Model CR2). The conditions for individual spectra are given in the captions. Further details of Raman instrumentation have been given elsewhere (Thomas, 1970).

Results and Conclusions

Spectra of Capsids and Disks. Secondary Structure of the Polypeptide Chain. Raman spectra of aqueous solutions of capsids and disks (Figure 1) show lines due to the normal modes of vibration of the amide groups which indicate the approximate molecular fractions of α-helical, β-sheet, and irregular structures present in protein molecules (Lord & Mendelsohn, 1972; Chen & Lord, 1974; Frushour & Koenig, 1974; Painter & Koenig, 1975; Lippert et al., 1976). The strong and sharp line from the amide I mode at 1655 cm^{-1} in the spectrum of the capsids or disks falls near the frequency characteristic of α helix and suggests that a significant fraction of the polypeptide chains in both the capsids and disks is in this configuration. A moderately intense line from the amide III mode is seen near 1248 cm^{-1} for both the capsids and disks, which indicates that a significant fraction of irregular structure is present in the coat protein molecules of these aggregates. The weak line near 1278 cm^{-1}, also an amide III mode, indicates α helix and confirms the conclusion drawn from the amide I mode. The absence of significant Raman intensity near 1235 cm^{-1} indicates that little if any β-sheet structure exists for TMV coat protein molecules when in the form of capsids or disks.

The line at 933 cm^{-1} arises from a skeletal C–C–N mode of the amide group and is also indicative of α-helical structure (Yu & Jo, 1973; Frushour & Koenig, 1974). The intensity of the line at 933 cm^{-1} appears to be significantly greater in the spectrum of disks than in the spectrum of capsids. For example, the intensity ratio I_{933}/I_{1005} is 0.31 ± 0.02 for the disks and 0.22 ± 0.02 for the capsids. Here, the choice of I_{1005} as the normalization basis is arbitrary. This may indicate substantially more α helix in disks than in capsids, but further experiments are needed to verify the validity of using the 1005-cm^{-1} line (due to phenylalanine residues) as a normalization basis. We note that in most (Frushour & Koenig, 1975) but not all (Thomas & Day, 1981) protein aggregates, the 1005-cm^{-1} line is relatively invariant to changes in state of aggregation. In addition, other amino acid residues may contribute to the Raman intensity observed at 933 cm^{-1}, so that the observed intensity difference between capsids and disks may have a more complex origin.

Two empirical methods have been described to estimate the fraction of polypeptide chains in the α-helical, β-sheet, and irregular structures by using amides I and III Raman intensities. The method of Lippert et al. (1976) requires spectra of deuterated protein, which are not available for TMV components. The recent method of Williams et al. (1980) is currently being evaluated for its applicability to TMV. Alternatively, we have compared the Raman spectra of the TMV capsids and disks with Raman spectra of capsids from turnip yellow mosaic virus (TYMV) (Hartman et al., 1978) and with the spectra of several proteins (Lippert et al., 1976) for which the secondary structures are known. By comparing the amide bands mentioned above, we arrive at the following estimate of secondary structure for protein subunits of TMV capsids and disks: 40–50% α helix, 40–50% irregular structure, and 0–20% β sheet. The imprecision in these estimates is the consequence of differences among the model proteins and

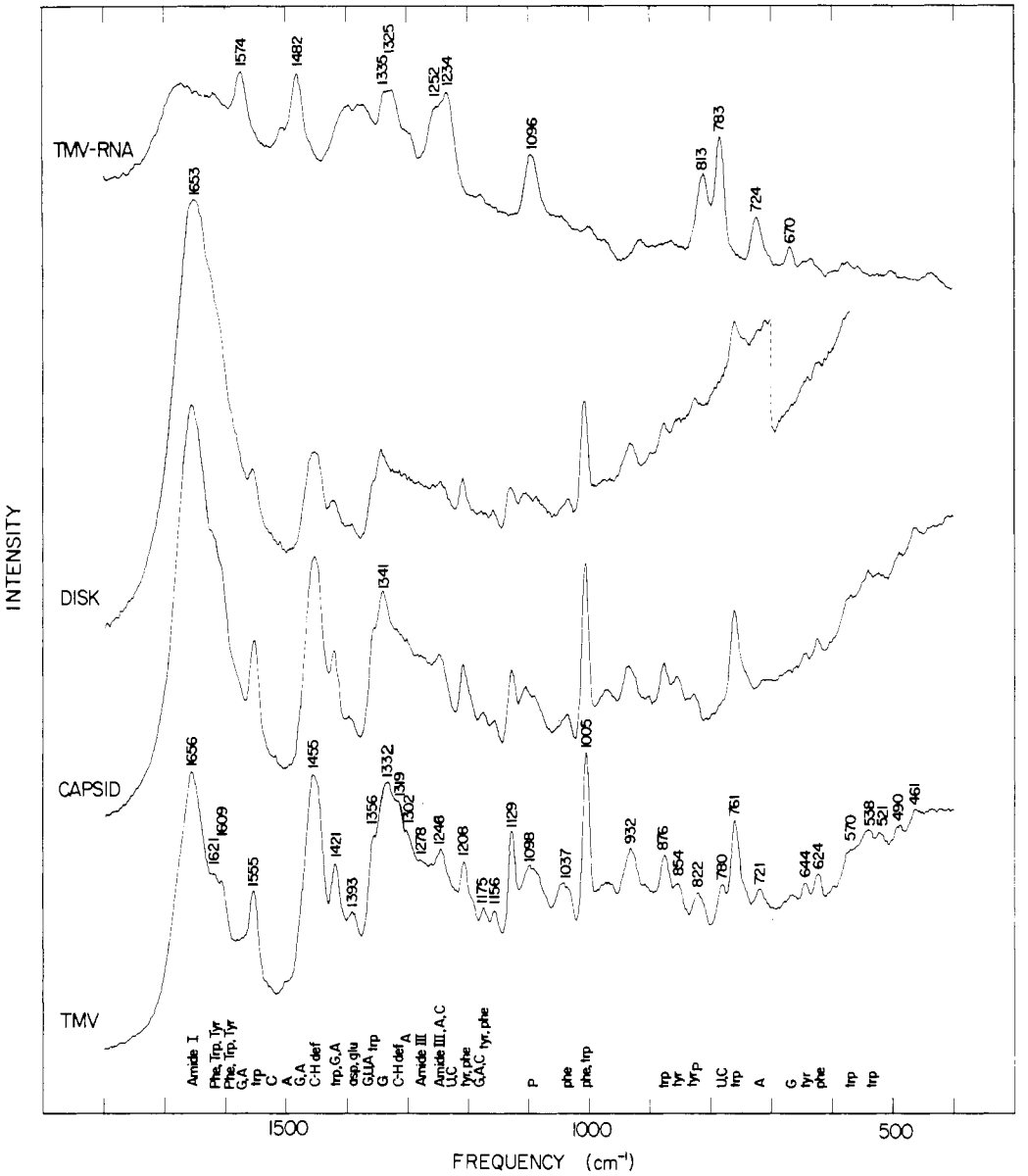


FIGURE 1: Raman spectra of TMV [7% (w/v) in 0.02 M Tris–NO_3, pH 7.0], capsid [6% (w/v) in distilled H_2O, pH 5.5], disk [4% (w/v) in 0.8 M KCl, pH 8.0], and TMV RNA [3% (w/v) in 1.0 mM NaCl, pH 7.0]. The sample temperature is 32 °C for all spectra with the exception of TMV RNA which is thermostated at 0 °C. Conditions: wavelength = 514.5 nm, power = 300 mW, slit width = 10 cm^{-1}, scan speed = 25 cm^{-1} min^{-1}, and rise time = 10 s.

uncertainties in base lines of the spectra.

Two other differences are observed between the spectra of capsids and disks which may be related to protein structure. First, the capsid has a Raman line at 975 cm^{-1}. This line is much weaker in the spectrum of the disk. Second, the region from 1070 to 1150 cm^{-1} shows clear differences when capsids and disks are compared. Again, the interpretation of these differences will require further work.

Environment of the Tryptophan Residues. The indole rings of the three tryptophan (Trp) residues known to exist in each TMV coat protein molecule give rise to a number of lines in the spectra of capsids and disks. Those near 1555 and 761 cm^{-1} are prominent and may be used as intensity references while those at 1356 and 877 cm^{-1} are sensitive to the environment of the indole rings (Frushour & Koenig, 1975; Kitagawa et al., 1979).

When the indole groups are surrounded by other hydrophobic groups, the line near 1356 cm^{-1} is strong and sharp and manifests an intensity comparable to the intensities of other Trp lines at 1555 and 761 cm^{-1} (Chen et al., 1974). For Trp residues exposed to solvent, the 1356-cm^{-1} line has an intensity value similar to the 1555 and 761 cm^{-1} lines and is partially resolved at moderate resolution (Figure 1). We conclude that at least two of the three Trp residues per coat protein molecule are in hydrophobic environments in both the capsid and disk structures.

Recent work suggests that the 877-cm^{-1} line may provide a quantitative estimate of the fraction of Trp groups in hydrophobic environments (Kitagawa et al., 1979). The ratio of the intensity of this line (I_{877}) to that of the line at 1455 cm^{-1} (I_{1455}) is said to be near 0.3 when all indole groups are in hydrophobic environments and near 0.1 when they are exposed to solvent (water). Within experimental uncertainty, we find I_{877}/I_{1455} is 0.23 for both capsid and disk, which indicates according to the correlation proposed by Kitagawa et al. (1979) that two of three indole groups per coat protein molecule exist in hydrophobic environments.

Environment of the Tyrosine Residues. The four tyrosine (Tyr) residues in each TMV coat protein molecule give rise to Raman lines near 854 and 827 cm^{-1} which are sensitive to the state of hydrogen bonding of the OH group of Tyr (Siamwiza et al., 1975). (These and other lines assigned to Tyr

Table I: Possible Distributions of Tyrosines among Different Bonding States in Disk, Capsid, and Virus[a]

states	no. of bonds		
	A	B	C
CV-1	0	4	0
CV-2	1	2	1
CV-3	2	0	2
D-1	1	0	3
D-2	0	2	2
D-3	0	3	1

[a] Abbreviations: CV denotes capsid or virion and D denotes disk. A, B, and C refer to hydrogen-bonding configurations of tyrosines (Siamwiza et al., 1975) as defined in the text.

are so labeled in Figure 1.) The ratio of intensities for these lines ($R_{Tyr} = I_{854}/I_{827}$) has been correlated with three kinds of hydrogen bonds. (A) For an OH group of Tyr which accepts a strong hydrogen bond from a positive group, e.g., from $-NH_3^+$ of lysine, R_{Tyr} is approximately 2.5. (B) For an OH group which forms moderate hydrogen bonds with water or with other similar OH groups, R_{Tyr} is near 1.25. (C) For an OH group which donates a strong hydrogen bond to a negative acceptor, e.g., to the carboxylate group of glutamic acid, R_{Tyr} is near 0.3.

Each of the four tyrosine residues in a protein subunit might participate in a type A, B, or C hydrogen bond, and each can contribute Raman intensity at 854 and 827 cm^{-1}. Therefore the value observed for R_{Tyr} represents the average from four Tyr residues per subunit. In principle, there exist 15 combinations of four indistinguishable groups distributed in three bonding situations with no restrictions as to the number with a given kind of bond. Restrictions may result, however, from tertiary structure which determines the groups available for hydrogen bonding with the Tyr residues (see Discussion). We have used the observed value of R_{Tyr} for capsid (1.31 ± 0.15) to eliminate some of the 15 alternative distributions, assuming equivalency of each protein molecule in the disk or capsid and assuming no restrictions from tertiary structure.

The data of Siamwiza et al. (1975) suggest the following empirical equation from the observed value of R_{Tyr}, to within 10% accuracy:

$$R_{Tyr} = 2.5X_A + 1.25X_B + 0.3X_C$$

where X_A, X_B, and X_C represent the fractions of Tyr residues participating in type A, B, and C bonds, respectively. On the basis of this equation, there are three possibilities consistent with the value of R_{Tyr} measured for capsids. These are shown in Table I. The twelve remaining combinations are excluded by the data.

For protein molecules in the form of the disk, R_{Tyr} was found to be 0.91 ± 0.13. Application of the above equation suggests three possible distributions of hydrogen bonding as shown in Table I. Note that all of these distributions are different from those proposed for the capsid.

Spectrum of TMV RNA. The Raman spectrum of RNA extracted from TMV (Figure 1) is quite similar to spectra of RNA from other viruses (Thomas et al., 1976; Turano et al., 1976) and of rRNA and tRNA from *E. coli* (Thomas, 1970; Thomas et al., 1980). Two of the lines in this spectrum which arise from the phosphodiester linkages have been used to estimate the amount and kind of secondary structure present in RNA (Lafleur et al., 1972). In this method, the intensity of the line near 1096 cm^{-1} (I_{1096}) serves as an intensity standard. A line near 814 cm^{-1} indicates the backbone structure characteristic of aqueous RNA and polyribonucleotides designated the "A RNA" structure (Arnott, 1970). Further, the intensity ratio $R_P = I_{814}/I_{1096}$ is a measure of the amount of such type-A structure present (Thomas & Hartman, 1973).

We measured R_P for TMV RNA solutions at 0 and 32 °C and found values of 1.47 ± 0.06 and 1.29 ± 0.02, respectively. These values indicate that approximately 90% of the RNA backbone has type-A structure at 0 °C, and this reduces to 79% at 32 °C. This result is similar to those obtained for most other kinds of RNA (Hartman et al., 1973; Thomas et al., 1976, 1980), but TMV RNA has more type-A structure than was found in TYMV RNA (Turano et al., 1976). The moderate decrease in type-A structure with increasing temperature over the range 0–32 °C has been observed for MS2 RNA (Thomas et al., 1976) and rRNA (Thomas et al., 1980) as well as for tRNA and 5S RNA (Chen et al., 1975, 1978).

The remaining strong lines in the RNA spectrum are assigned to ring vibrational modes of the bases. Of these, the lines at 670 and 1482 cm^{-1} are hyperchromic; i.e., the intensity increases as the bases stack in a regular structure, while most of the other Raman lines are hypochromic, i.e., the intensity decreases with base stacking (Small & Peticolas, 1971a,b; Lafleur et al., 1972; Morikawa et al., 1973; Prescott et al., 1974). Since, in TMV RNA at 0 °C, 90% of the bases participate in type-A structure, extensive stacking and pairing of bases is possible, and appropriate hyperchromic and hypochromic effects should exist in the spectrum. Thus, the normally weak line from guanine residues at 670 cm^{-1} is relatively strong in the spectrum of Figure 1 as expected for guanine in a stacked structure (hyperchromism), and the normally well-resolved line of cytosine near 1530 cm^{-1} has been reduced to a nondetectable level by stacking of the cytosines (hypochromism). [The mole fractions of guanine and cytosine in TMV RNA are 0.24 and 0.20 (Knight, 1975), respectively, which should also be taken into account when the data of Figure 1 are compared with those of other RNAs.]

Spectrum of the Virus. Secondary Structure of the Polypeptide Chain. The Raman spectrum of an oriented gel of TMV, prepared according to the method of Gregory & Holmes (1965), shows a strong line at 1656 cm^{-1}, due predominantly to the amide I mode of coat protein molecules (Figure 1). The relatively weak scattering from RNA near 1650–1700 cm^{-1} does not influence the position of the amide I peak. Amide I of TMV is therefore positioned at essentially the same frequency found for the capsids and disks and suggests that the coat protein molecules in the virus have an amount of α helix similar to that in the subunits of capsids and disks. This conclusion is supported by the intensity of the line near 933 cm^{-1} from the C–C–N skeletal mode for which $I_{933}/I_{1005} = 0.31 \pm 0.02$.

In the amide III region of TMV, it is difficult to interpret quantitatively the significance of the Raman line near 1246 cm^{-1}, since the observed intensity here includes small contributions from adenine, cytosine, and uracil residues of the RNA (Thomas, 1970). Nevertheless, the overall band contours are similar to those observed for the capsids and disks, which supports the conclusion that coat protein molecules of the virion have approximately the same amounts of α-helical, β-sheet, and irregular structures as in the capsid and disk states.

Environment of the Tryptophan Residues. As for capsids and disks, the Raman spectrum of TMV shows a discernible shoulder at 1356 cm^{-1}. The broad scattering envelope in this region expected from RNA does not obscure this feature. A high-resolution scan of this region (Figure 2) clearly reveals

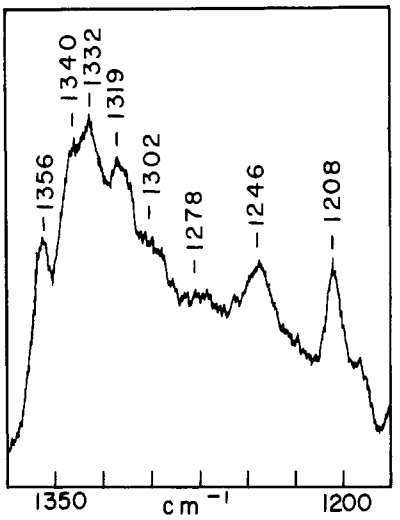


FIGURE 2: High-resolution Raman spectrum of TMV [7% (w/v) in 0.02 M Tris–NO_3, pH 7.0]. The sample temperature is 32 °C. Conditions: wavelength = 514.5 nm, power = 300 mW, slit width = 4 cm^{-1}, scan speed = 25 cm^{-1} min^{-1}, and rise time = 3 s.

the line of Trp at 1356 cm^{-1}. The intensity of the 1356-cm^{-1} line relative to lines at 761 and 1555 cm^{-1} again suggests that the majority of Trp residues exist in hydrophobic environments as found for capsids and disks.

The line from Trp at 877 cm^{-1} is separated from scattering of RNA, and its intensity may be reliably measured as was done for capsids and disks. We find the ratio I_{877}/I_{1455} to be 0.23, which is the same as for capsids and disks and which supports the conclusion that at least two of three Trp residues per coat protein molecule exist in hydrophobic environments (Kitagawa et al., 1979).

Environment of the Tyrosine Residues. The region 800–850 cm^{-1} of the Raman spectrum of the virion is expected to contain lines from vibrational modes of the RNA backbone (Thomas & Hartman, 1973) and from the tyrosine residues of the protein (Siamwiza et al., 1975). As seen in Figure 1, the lower frequency line of the tyrosine doublet (expected near 828 cm^{-1}) and the RNA line (expected near 815 cm^{-1}) are fused to give the appearance of a single broad line centered at approximately 822 cm^{-1}.

The situation is clarified by a high-resolution spectral scan of this region (Figure 3) which reveals three distinct lines at 812, 820, and 827 cm^{-1}. Although this complexity precludes the simple computation of R_{Tyr} as done for capsids and disks, we have estimated this value by an indirect approach. We assign with confidence the lines at 812 and 820 cm^{-1} to RNA (see the following section) and the line at 827 cm^{-1} to Tyr. The RNA lines actually give some residual intensity at 827 cm^{-1} as well as at lower frequency values where Tyr would no longer scatter, which makes the choice of base line uncertain. To circumvent this problem, we proceeded as follows. First, we estimated the scattering envelope of the Trp line at 760 cm^{-1}, extended to its apparent minimum near 814 cm^{-1} as deduced from the spectrum of the capsid, Figure 1. This is justified since Trp residues occur in equal numbers in capsid and virion and the line at 760 cm^{-1} is independent of environment (Kitagawa et al., 1979). Next a straight line was drawn from the minimum of the estimated scattering envelope at 814 cm^{-1} to the next minimum in the spectrum at about 900 cm^{-1}. The resulting tangent was used as a base line to measure I_{855} and I_{827}, from which R_{Tyr} was computed to be ~1.3. This suggests that the bonding of the four OH groups of the tyrosine residues in each coat protein molecule has the same three possibilities in the virion as determined for the

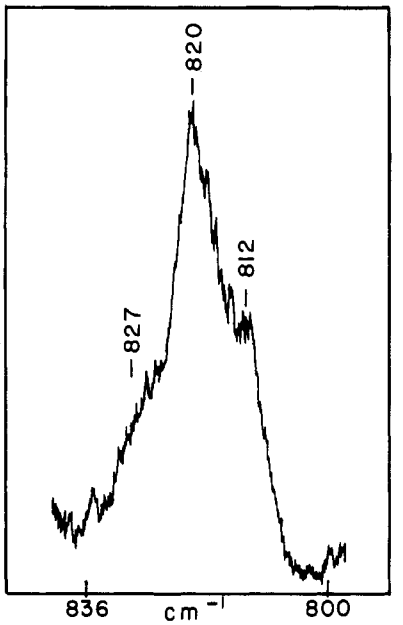


FIGURE 3: High-resolution Raman spectrum of TMV [5% (w/v) in 0.02 M Tris–NO_3, pH 7.0]. The sample temperature is 32 °C. Conditions: wavelength = 514.5 nm, power = 300 mW, slit width = 4 cm^{-1}, scan speed = 10 cm^{-1} min^{-1}, and rise time = 10 s.

capsid (Table I), but the certainty of this conclusion is obviously lower for virions, because of the assumptions made in the above described procedure.

Conformation of RNA in the Virion. The virion of TMV contains only 5 wt % RNA, which suggests that Raman lines from RNA in the spectrum of the virion will be weak and often obscured by protein. This is indeed the case for the 1650–1700-cm^{-1} region of the spectrum discussed above. In searching for other RNA lines, we compared numerous spectral scans of the virion with scans of protein (disks and capsids) and RNA, using varying conditions of spectral resolution. This allowed the identification of the following RNA lines in the spectrum of the virion.

The normally strong RNA line at 1574 cm^{-1} (due to adenine and guanine) is seen as a "filling-in" of the minimum observed near this frequency in the spectrum of the capsid. The equally strong line due to adenine and guanine at 1482 cm^{-1} in the RNA spectrum is not resolved in the spectrum of the virion shown in Figure 1 but is clearly visible as a weak shoulder near 1480 cm^{-1} in a high-resolution spectrum of the virion (not shown in Figures 1–3). The RNA lines near 1325 and 1335 cm^{-1} contribute to the broad and intense Raman scattering observed in this region for the virion. The strong RNA lines due to cytosine and uracil near 1240 cm^{-1} are observed as increased intensity in the amide III region of the spectrum of the virion. The line near 1096 cm^{-1} due to PO_2^- groups of RNA is difficult to locate on top of the broader band from protein in this region. The examination of many spectra suggests only a slight increase in intensity at 1096 cm^{-1} which is not readily apparent in Figure 1. The important RNA line expected near 814 cm^{-1} (Thomas, 1970) is seen instead as part of the composite band at 822 cm^{-1} and is resolvable into two RNA components at 820 and 812 cm^{-1} (Figure 3). The overlapping and strong lines due to uracil and cytosine (780 cm^{-1}) as well as the moderately intense line due to adenine (721 cm^{-1}) are prominent features in the spectrum of the virion. The guanine line at 670 cm^{-1} is weak but is always observed in TMV.

As stated in the preceding paragraph, in the spectrum of TMV only very little intensity due to RNA is seen at 1096 cm^{-1}. Although I_{1096} cannot be measured precisely, it is clear

in Figure 1 that I_{721}/I_{1096} is much greater than unity in the virion whereas the corresponding ratio in RNA (I_{724}/I_{1096}) is less than unity. Since the adenine line at 721 cm^{-1} is known to be hypochromic, this result suggests that its intensity is enhanced in TMV vis-ā-vis purified TMV RNA; i.e., in TMV most adenine bases are no longer stacked in the same way that occurs in protein-free RNA. The same argument may be applied to the hypochromic lines at 1574 (A and G), 1246 (U and C), and 783 cm^{-1} (C and U) and to the complex of lines between 1295 and 1350 cm^{-1}. For each of these lines, I/I_{1096} is larger in TMV than in protein-free TMV RNA.

It is therefore clear from numerous lines in the Raman spectra that the RNA bases have a very different arrangement in the virion than in the protein-free state. This difference is consistently in favor of less base stacking in the virion. We turn now to an analysis of the Raman lines of the RNA backbone.

At high resolution (Figure 3), the spectrum of TMV has a clear-cut peak at 820 cm^{-1} with a well-resolved shoulder at 812 cm^{-1} and a diffuse shoulder at 827 cm^{-1}. Since spectra of the capsids and disks have a line due to Tyr near 828 cm^{-1}, we conclude that the shoulder located at 827 cm^{-1} in the spectrum of the virion is due to Tyr. Furthermore, since neither the disks nor capsids have Raman lines near 812 or 820 cm^{-1}, we assign these lines to RNA. It is evident also from Figure 3 that the 820-cm^{-1} component is about twice as intense as the 812-cm^{-1} feature.

If these assignments are correct, they indicate a significant change in the phosphodiester group (C–O–P–O–C) frequency for RNA in the virion, as compared with protein-free aqueous TMV RNA. One interpretation of these data is the following. At least two of the three RNA nucleotides which bind to each TMV protein subunit, by virtue of their unusual phosphodiester frequency (820 cm^{-1}), have a substantially different geometry than has been previously observed by Raman spectroscopy for aqueous RNA and polyribonucleotides. The third nucleotide, on the basis of its phosphodiester frequency (812 cm^{-1}) could have a geometry similar to that normally occurring in aqueous or crystalline RNA (Thomas, 1970; Chen et al., 1975), polyribonucleotides (Thomas & Hartman, 1973), and A DNA (Erfurth et al., 1972).

On the other hand, recent results from X-ray diffraction analysis (Stubbs & Stauffacher, 1981) suggest that although all three nucleotides of the TMV binding site have the C3′-endo pucker of the ribose ring characteristic of A DNA and A RNA (Arnott, 1981), none of them has the g^-g^- conformation of the C–O–P–O–C dihedral angles generally found in helical RNA structures. Thus, X-ray diffraction data indicate that *all three nucleotides of the binding site differ conformationally from the classical type-A conformation.* These TMV conformers have been given as tg^+, g^+g^+ and g^+g^- (Stubbs & Stauffacher, 1981).

If the structure of TMV in crystalline (X-ray) and aqueous (Raman) states is the same, then we must conclude that two of the nucleotide conformers yield essentially the same Raman phosphodiester frequency (820 cm^{-1}) and the third yields a different frequency (812 cm^{-1}) that happens to nearly coincide with the frequency of the g^-g^- conformer. Further work on model compounds will be required to determine which two of the three conformers are responsible for the 820-cm^{-1} frequency.

Discussion

A number of problems slowed the work reported here. In 1975, the small quantity of TMV on hand limited the concentration of samples for Raman spectroscopy, causing weak lines with poor signal to noise. This was remedied by the gift of significant quantities of TMV from Dr. C. A. Knight. Concentrated solutions (up to 100 mg/mL) were later prepared which gave much better spectra, but one major problem remained. The scattering background fluctuated slowly which caused the measured intensities of supposedly stable lines (e.g., I_{1455} and I_{1055}) to vary greatly from scan to scan. This prevented quantitative analysis of the tyrosine and tryptophan lines given above and precluded reliable interpretation of the data.

The fluctuation of the background intensity appears to be due to nonspecific aggregation of the virions in concentrated solutions which could not be prevented by varying the ionic strength and pH. The problem was finally eliminated by the use of oriented gels of TMV (Gregory & Holmes, 1965) from samples donated by Dr. Holmes. These samples are similar to those used in X-ray diffraction studies. The Raman spectra of virus from both sources were identical except for the fluctuating background of the former.

In a recent Raman study (Fox et al., 1979), spectra have been described as those of TMV, its RNA, and A protein. However, all of the spectra of Fox et al. (1979) are very different from those obtained by us. Their sample of A protein appears to be much less concentrated than 10% protein in water and surprisingly contains no characteristic Raman lines of Trp at 1555, 1424, 1368, and 878 cm^{-1}. A similar problem applies to their spectrum of TMV, which could be contaminated with other Raman active impurities. Their spectrum of TMV RNA contains lines at 1460, 1126, 1180, 1050, 998, and 888 cm^{-1} which have intensities not typical of RNA (Thomas, 1970) and might be due in part to inorganic phosphate contaminants (Chapman & Thirlwell, 1964).

Shie et al. (1978) have also published and interpreted a Raman spectrum of TMV in aqueous solution. Their results and conclusions are similar to ours except as noted below.

A number of weak maxima and shoulders (e.g., at 1718, 1706, 1666, 1405, 1085, 980, 816, and 796 cm^{-1}) are found in their spectrum but not in ours. One explanation might be that they used a higher resolution than that stated (say 5 cm^{-1} instead of 10 cm^{-1}). A second more plausible explanation is that the Raman lines are affected by contaminants, such as the 0.01 M phosphate buffer which they employed.

The Raman frequencies cited by Shie et al. (1978) at 816 (peak) and 796 cm^{-1} (shoulder) are very important since they are in the region of conformationally sensitive, phosphodiester frequencies discussed above. We have reviewed all of our spectra of TMV taken both before and after the publication of the data of Shie et al. (1978), at resolutions of 5 cm^{-1} as well as 10 cm^{-1}, and in no case do we find a shoulder at 796 cm^{-1}. The latter value is in fact near a minimum in our spectra. Therefore we do not agree with the conclusions of Shie et al. (1978) regarding both the interpretation of the structure of RNA in TMV and the significance of the Raman phosphodiester frequency.

Another major difference in the spectra is the intensity ratio of the tyrosine lines, R_{Tyr}. Shie et al. obtained $R_{Tyr} = 0.75$, a value much lower than the highly reproducible value ($R_{Tyr} = 1.31 \pm 0.15$) obtained by us. We note that our value was computed after correction for intensity due to RNA. Shie et al. (1978) make no mention of such a correction.

Regarding assignments of the lines in the Raman spectrum of the virus, Shie et al. (1978) assign lines above 1700 cm^{-1} to $-CO_2H$ groups and lines near 1390 and 1415 cm^{-1} to $-CO_2^-$ groups from aspartic and glutamic acid residues. We believe these assignments are unlikely since the relevant vibrational

modes give very weak scattering, and at pH 7.2, only one carboxyl group per coat protein molecule would be protonated (Scheele & Lauffer, 1967; Butler et al., 1972). Furthermore, the lines at 1390 and 1415 cm^{-1} do not diminish in intensity, and the lines above 1700 cm^{-1} do not increase in intensity as the pH is lowered to 5.5 to titrate the second carboxyl group (K. A. Hartman, P. E. McDonald-Ordzie, B. Prescott, and G. J. Thomas, Jr., unpublished data). Such changes would be expected if $-CO_2^-$ groups were protonating to form $-CO_2H$ groups and are not observed in the Raman spectra of β-lactoglobulin recorded vs. pH in the range 8.0–4.0 (S. R. Fish, K. A. Hartman, and G. J. Thomas, Jr., unpublished data).

In this work we have estimated that coat protein molecules in the capsid, disk, and virus have 40–50% of their peptide groups in α-helical regions, 40–50% in irregular structures, and 0–20% in the β-sheet structure. These values agree with those derived from X-ray studies.

Bloomer et al. (1978) have obtained electron density maps of the disk at a resolution of 2.8 Å and have located the amino acid residues with high accuracy. The four regions of α helix in the disk contain 60 residues or 38% of the coat protein molecule. Residues 49–52 comprise a 3_{10} helix for a total of 64 residues or 40.5%. In addition, 13 residues participate in a β-sheet structure (i.e., 8.2%). This would leave about 51% of the amino acid residues in what we have called irregular structures which connect the regions of α helix and β sheet.

Stubbs et al. (1977) have obtained electron density maps of the virus at a resolution of 4 Å which allowed them to fit idealized α-helical structures. They assigned four α-helical regions which are similar to those in the disk (Champness et al., 1976) and an additional region of α helix (called "vertical") at smaller radius for a total of 74 residues or 46.8% of the coat protein molecule. Those portions of the molecule farther than 65 Å from the axis of the particle were not as well resolved so that regions of β sheet and possibly other helical regions as well were not observed by Stubbs et al. (1977). However, X-ray diffraction data do suggest that protein molecules have very nearly the same structure in the capsid, disk, and virus (Stubbs et al., 1979; Bloomer et al., 1978).

We have measured the intensity of the 933-cm^{-1} line normalized to the intensity of the 1005-cm^{-1} line in spectra of the capsid, disk, and virion and find the capsid to have less intensity at 933 cm^{-1} than either disk or virion. This could be interpreted (Frushour & Koenig, 1975) to suggest that the capsid has less α helix than the virion or disk, which does not agree with X-ray results. However, no quantitative method has been proposed for the evaluation of α helix from the 933-cm^{-1} line intensity, and other lines from the C–C–N network may also be involved. For instance, the capsid has a well-defined moderate peak at 973 cm^{-1} which is missing from the spectrum of the disk. Reliable interpretation of these lines, which have varying frequencies from protein to protein (Thomas & Day, 1981; Thomas et al., 1981), must await further study, but if the intensity of the 933-cm^{-1} line is proportional to the fraction of α helix in the TMV protein molecule, we would expect higher resolution diffraction analysis to show that the capsid contains fewer helical regions than exist for the disk and virion.

A model for the RNA molecule in the TMV particle has also been derived from diffraction data (Stubbs et al., 1977; Stubbs & Stauffacher, 1980, 1981). In this model, the bases reside against hydrophobic regions of the left radial helix which precludes the possibility of base-stacking interactions. Also, the C–O–P–O–C conformations of the three nucleotides associated with the protein binding site are significantly different from one another (tg^+, g^+g^+, and g^+g^-) and from the classical A conformation (g^-g^-) (Stubbs & Stauffacher, 1981). Our conclusions on the RNA backbone in TMV agree with the X-ray results, by showing that at least two of the three backbone linkages are geometrically different from those normally occuring in aqueous RNA or A DNA. The Raman picture of RNA in the virus particle is thus quite different from that in protein-free RNA. Moreover, in the virion, the hypochromism of many lines due to the RNA bases has been lost, which conclusively shows that the bases no longer stack as they do in protein-free RNA, again in accord with the X-ray model.

The present results on RNA phosphodiester group frequencies in TMV support the growing recognition that nucleic acid structures do not fall into a few narrow groups, such as the A, B, and C forms of DNA. Instead, nucleic acid structures occupy diverse positions in a wide range of possible conformations. This has been pointed out by Arnott (1981) for helical polynucletides and should be even more the case for protein-associated viral nucleic acids. The appearance of a phosphodiester group frequency at 820 cm^{-1} in TMV thus signifies a new RNA conformation detectable by Raman spectroscopy. Further study of model compounds will be required to answer the important questions of which conformers yield this frequency and to what extent are the phosphodiester group frequencies perturbed by intermolecular interactions involving H_2O, metal ions, and amino acid side chains.

The Raman spectra of TMV and its capsid and disk have provided estimates of the environment and bonding of tryptophan and tyrosine residues which may be compared with models of the coat protein molecule derived from X-ray diffraction studies. The model derived for the disk is clear at radii greater than 60 Å, where all tyrosine and tryptophan residues are located (Bloomer et al., 1978). The model for the virus (Stubbs et al., 1977) does not show this region, but comparisons suggest that details are the same in virus, capsid, and disk (Bloomer et al., 1978; Stubbs et al., 1979). Fluorescence spectroscopic studies also suggest that the environment of tryptophan and tyrosine residues is not changed among monomer, A protein, and capsid (Guttenplan & Calvin, 1973), although work using circular dichroism predicts that some changes in tyrosine environment do occur during aggregation of protein. [For a review of chemical and spectroscopic studies, see Kaper (1975).] In the following, we will assume that the model for the disk also gives the approximate locations and environments for the capsid and virus. Evidence from chemical studies will also be mentioned.

Raman spectra predict that two out of three tryptophan residues are in hydrophobic environments and that one is in contact with solvent water. This is in good agreement with the model of X-ray work and with the results of chemical studies. The former places Trp-17 in the β-sheet structure and Trp-52 between the β sheet and the right slew α helix. Both are in a hydrophobic region of the protein molecule. Trp-152 is six residues from the carboxyl terminus and is located on the outer surface of the molecule (at a radius of ~85 Å), which suggest contact with water. Hydrolysis with carboxypeptidase shows that up to 22 of the final amino acid residues are on the outside of the molecule in the virus, and this is supported by serological and chemical studies (Kaper, 1975). Although Trp-152 does not react with *N*-bromosuccinimide, this is probably due to an unfavorable orientation which excludes the reagent rather than to a complete screening of the indole side group by other hydrophobic residues.

The Raman spectra of the virus and capsid predict three possible distributions of the four Tyr residues among three kinds of hydrogen bonds (see Table I). The model of the protein molecule from X-ray diffraction of the disk shows that none of the Tyr residues is hydrogen bonded to either lysine or arginine residues (type-A bonds) or glutamic or aspartic acid residues (type-C bonds), and it is very likely that this applies to the capsid and virus also.

The second residue from the amino terminus is on the surface of the model and is chemically reactive under certain circumstances. The *p*-hydroxyphenyl group could project toward the hydrophobic region at smaller radius which could sterically limit reactivity, but the total exclusion of water from the OH group seems unlikely. Therefore Tyr-2 should form type-B hydrogen bonds with water.

Tyr-70 is in the β-sheet structure with its side chain distal to the sheet pointing out of the hydrophobic region. On the basis of the published ribbon model, which does not show the positions of side groups, it is likely that the OH group of Tyr-70 also forms type-B bonds with water.

Tyr-72 is between the β sheet and the right radial α helix and appears to project into the space between protein molecules. From the model presented, we cannot judge the accessibility to water of Tyr-72.

Tyr-139 is in a hydrophobic cluster with Tyr-70 and Trp-17 but is considered to be the most chemically reactive of the four tyrosine residues, which suggests that water molecules could approach the OH group of Tyr-139.

The X-ray results eliminate the second two alternative distributions given in Table I (i.e., CV-2 and CV-3). Therefore, Raman and X-ray data considered together suggest that all Tyr residues participate in moderate hydrogen bonds with water molecules (or the OH groups of Ser or Thr residues) in the capsid and virus.

Although we expect the same result for the disk as for the capsid and virus, the Raman spectra predict three very different distributions (D-1, D-2, and D-3 in Table I), one of which requires one type-A bond and the others require combinations of type-B and type-C bonds. Neither type-A nor type-C bonds are likely from the X-ray models. It appears that a conflict exists between the Raman data and the results from X-ray studies. The resolution probably lies in the fact that we prepared the samples of disk by increasing the concentration of KCl to 0.8 M at pH 8 (rather than titrating samples of small volume into a narrow pH region). Chloride ions may accept strong hydrogen bonds and might substitute for $-CO_2^-$ in providing type-C bonds for tyrosine in our samples of disk. (We plan to check this by preparing disks in low concentrations of KCl as were used with capsids in the present work.) If this explanation is correct, we may eliminate the first two alternatives for the disk (see Table I) and state that three type-B bonds (to H_2O or similar OH groups from serine or threonine) and one type-C bond (to Cl^-) exist in disks prepared by our method. This, if correct, would demonstrate that one of the tyrosine residues is so located as to allow the approach of Cl^- while three are not.

References

Arnott, S. (1970) *Prog. Biophys. Mol. Biol. 21*, 265–319.

Arnott, S. (1981) *Trends Biochem. Sci. (Pers. Ed.) 5*, 231–234.

Bakke, M. K. (1979) *Virology 98*, 76–87.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., & Klug, A. (1978) *Nature (London) 276*, 362–368.

Butler, P. J. G., Durham, A. C. H., & Klug, A. (1972) *J. Mol. Biol. 72*, 1–18.

Caspar, D. L. D. (1963) *Adv. Protein Chem. 18*, 37–121.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G., & Klug, A. (1976) *Nature (London) 259*, 20–24.

Chapman, A. C., & Thirlwell, L. E. (1964) *Spectrochim. Acta 20*, 937–947.

Chen, M. C., & Lord, R. C. (1974) *J. Am. Chem. Soc. 96*, 4750–4752.

Chen, M. C., Lord, R. C., & Mendelsohn, R. (1974) *J. Am. Chem. Soc. 96*, 3038–3042.

Chen, M. C., Giege, R., Lord, R. C., & Rich, A. (1975) *Biochemistry 14*, 4385–4391.

Chen, M. C., Giege, R., Lord, R. C., & Rich, A. (1978) *Biochemistry 17*, 3134–3138.

Durham, A. C. H. (1972) *J. Mol. Biol. 67*, 289–305.

Durham, A. C. H., & Finch, J. T. (1972) *J. Mol. Biol. 67*, 307–314.

Erfurth, S. C., Kiser, E. J., & Peticolas, W. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 938–941.

Fox, J. W., Lee, J., Amorese, D., & Tu, A. T. (1979) *J. Appl. Biochem. 1*, 336–343.

Fraenkel-Conrat, H. (1957) *Virology 4*, 1–4.

Frushour, B. G., & Koenig, J. L. (1974) *Biopolymers 13*, 1809–1819.

Frushour, B. G., & Koenig, J. L. (1975) in *Advances in Infrared and Raman Spectroscopy* (Clark, R. J. H., & Hester, R. E., Eds.) Heyden & Son, New York.

Gregory, J., & Holmes, K. C. (1965) *J. Mol. Biol. 13*, 796–801.

Guttenplan, J. B., & Calvin M. (1973) *Biochim. Biophys. Acta 322*, 294–310.

Hartman, K. A., Lord, R. C., & Thomas, G. J., Jr. (1973) *Phys. Chem. Prop. Nucleic Acids 2*, 1–89.

Hartman, K. A., McDonald-Ordzie, P. E., Kaper, J. M., Prescott, B., & Thomas, G. J., Jr. (1978) *Biochemistry 17*, 2118–2123.

Holmes, K. C. (1980) *Trends Biochem. Sci. (Pers. Ed.) 5*, 4–7.

Kaper, J. M. (1975) *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, North-Holland Publishing Co., New York.

Kitagawa, T., Azuma, T., & Hamaguchi, K. (1979) *Biopolymers 18*, 451–465.

Knight, C. A. (1975) *Chemistry of Viruses*, 2nd ed., Springer-Verlag, New York.

Lafleur, L., Rice, J., & Thomas, G. J., Jr. (1972) *Biopolymers 11*, 2423–2437.

Lippert, J. L., Tyminski, D., & Desmeules, P. J. (1976) *J. Am. Chem. Soc. 98*, 7075–7080.

Lord, R. C., & Mendelsohn, R. (1972) *J. Am. Chem. Soc. 94*, 2133–2134.

Morikawa, K., Tsuboi, M., Takahashi, S., Kyogoku, Y., Mitsui, Y., Iitaka, Y., & Thomas, G. J., Jr. (1973) *Biopolymers 12*, 799–812.

Painter, P. C., & Koenig, J. L. (1975) *Biopolymers 14*, 457–468.

Prescott, B., Gamache, R., Livramento, J., & Thomas, G. J., Jr. (1974) *Biopolymers 13*, 1821–1845.

Scheele, R. B., & Lauffer, M. A. (1967) *Biochemistry 6*, 3076–3081.

Shie, M., Dobrov, E. N., & Tikchonenko, T. I. (1978) *Biochem. Biophys. Res. Commun. 81*, 907–914.

Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., & Shimanouchi, T. (1975) *Biochemistry 14*, 4870–4876.

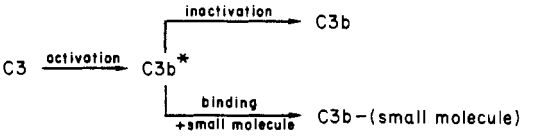
Small, E. W., & Peticolas, W. L. (1971a) *Biopolymers 10*, 69–88.

Small, E. W., & Peticolas, W. L. (1971b) *Biopolymers 10*, 1377–1418.
Stubbs, G., & Stauffacher, C. (1980) *Biophys. J. 32*, 244–245.
Stubbs, G., & Stauffacher, C. (1981) *J. Mol. Biol. 152*, 387–396.
Stubbs, G., Warren, S., & Holmes, K. (1977) *Nature (London) 267*, 216–221.
Stubbs, G., Warren, S., & Mandelkow, E. (1979) *J. Supramol. Struct. 12*, 177–183.
Thomas, G. J., Jr. (1970) *Biochim. Biophys. Acta 213*, 417–423.
Thomas, G. J., Jr. (1978) *Proc. Int. Conf. Raman Spectrosc., 6th 1*, 91–103.
Thomas, G. J., Jr., & Hartman, K. A. (1973) *Biochim. Biophys. Acta 312*, 311–322.
Thomas, G. J., Jr., & Day, L. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2962–2966.
Thomas, G. J., Jr., Prescott, B., McDonald-Ordzie, P. E., & Hartman, K. A. (1976) *J. Mol. Biol. 102*, 103–124.
Thomas, G. J., Jr., Prescott, B., & Hamilton, M. G. (1980) *Biochemistry 19*, 3604–3613.
Thomas, G. J., Jr., Prescott, B., Day, L. A., & Boyle, P. D. (1981) *Prog. Clin. Biol. Res. 64*, 429–440.
Turano, T. A., Hartman, K. A., & Thomas, G. J., Jr. (1976) *J. Phys. Chem. 80*, 1157–1163.
Williams, R. W., Dunker, A. K., & Peticolas, W. L. (1980) *Biophys. J. 32*, 232–234.
Yu, N.-T., & Jo, B. H. (1973) *J. Am. Chem. Soc. 95*, 5033–5037.

Binding Reaction between the Third Human Complement Protein and Small Molecules†

Sai-Kit A. Law,* Tana M. Minich, and R. P. Levine

ABSTRACT: The covalent binding reaction of the third complement protein (C3) to receptive surfaces is thought to proceed by the following mechanism. An internal thioester [Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L., & Prahl, J. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5764–5768; Law, S. K., Lichtenberg, N. A., & Levine, R. P. (1980b) *Proc. Natl. Acad. Sci. U.S.A. 77*, 7194–7198], which is usually hidden within the C3 molecule, is exposed upon proteolytic activation of C3 to C3b* (the hypothetical conformation of C3b which has the capacity to bind to receptive surfaces and small molecules). The exposed thioester is accessible to attack by hydroxyl groups on receptive surfaces. An acyl transfer reaction takes place, leading to the binding of C3b to the receptive surfaces via an ester linkage [Law, S. K., Lichtenberg, N. A., & Levine, R. P. (1979) *J. Immunol. 123*, 1388–1394]. We have used a fluid-phase system to demonstrate the specific binding of different small molecules to the labile binding site of C3. The small molecules include glycerol, different hexose monomers, sucrose, raffinose, and four amino acids. These molecules bind to C3b with different efficiencies, indicating that there is an order of preference of C3b* for these molecules. In certain cases, the small molecules bind to C3b via ester linkages (e.g., glucose); in others, the bond is an amide linkage (e.g., lysine). We have also studied the concentration dependence of the binding of small molecules to C3b. The binding is consistent with the following reaction scheme:



The binding of C3b, the activated form of the third complement protein (C3), to cell surfaces is a central step in the humoral immune defense against infection. The binding reaction between C3/C3b and receptive surfaces (RS)[1] has been studied by several laboratories, and a model of the chemistry of the binding reaction has been proposed. This model is schematically shown in Figure 1. There is an internal thioester within C3 (Janatova et al., 1980a; Pangburn & Müller-Eberhard, 1980; Howard, 1980; Tack et al., 1980; Law et al., 1980b) between a cysteinyl residue and a glutamyl residue (Tack et al., 1980). These amino acids are separated by two amino acid residues along the α chain of C3 (Tack et al., 1980). The thioester is probably protected in a hydrophobic pocket since it is quite stable in aqueous medium (Pangburn & Müller-Eberhard, 1980; Law et al., 1980b), and it is more accessible to methylamine than similar nucleophiles of progressively larger sizes such as ethylamine, isopropylamine, and *tert*-butylamine (Pangburn & Müller-Eberhard, 1980). The activation of C3 is initiated by the proteolytic conversion of C3 to C3a and C3b. Activation imparts to the C3b portion of the molecule the capacity to bind to receptive surfaces, and it has been described as the exposure of a labile binding site of C3 (Müller-Eberhard et al., 1966) that has a lifetime on

† From the James S. McDonnell Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110. *Received June 29, 1981.* This work was supported by National Institutes of Health Grant AI 16543.

* Address correspondence to this author at the Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom.

[1] Abbreviations: C3b*, the hypothetical conformation of C3b which has the capacity to bind to receptive surfaces and small molecules; RS, receptive surfaces; CH_3NH_2, methylamine; SBTI, soybean trypsin inhibitor; PMSF, phenylmethanesulfonyl fluoride; $NaDodSO_4$, sodium dodecyl sulfate; BE, binding efficiency; k, rate of conversion of C3 to C3b* by trypsin; k_0, rate of decay of C3b* to C3b in the fluid phase; k', rate of binding of glycerol to C3b* to form C3b–glycerol.

0006-2960/81/0420-7457$01.25/0