Phage φX174 Probed by Laser Raman Spectroscopy: Evidence for Capsid-Imposed Constraint on DNA Secondary Structure†

N. L. Incardona
Department of Microbiology and Immunology, University of Tennessee—Memphis, Memphis, Tennessee 38163
B. Prescott, D. Sargent, O. P. Lamba, and G. J. Thomas, Jr.*
Department of Chemistry, Southeastern Massachusetts University, North Dartmouth, Massachusetts 02747
Received August 11, 1986; Revised Manuscript Received November 14, 1986

ABSTRACT: The Raman spectrum of the isometric bacteriophage φX174 contains a number of well-resolved bands which have been assigned unambiguously to proteins of the capsid or to the single-stranded DNA (ssDNA) genome. Additional Raman bands of protein and DNA, which are partially overlapped in the spectrum of virus, have been resolution enhanced by Fourier deconvolution to permit improved semiquantitative measurement of spectral intensities and frequencies for structural conclusions. Raman conformation markers indicate that the ssDNA molecule within the capsid contains nucleosides of C2′-endo sugar pucker and *anti*-glycoside bond orientation, but the nucleic acid backbone lacks the geometry characteristic of B-form DNA. The Raman profile of encapsidated φX DNA indicates a backbone more similar to heat-denatured DNA than to DNA containing hairpinlike secondary structure. This finding suggests limited interbase interactions in the packaged genome, which is presumably the result of constraints imposed by the viral capsid. Thus, the extensive pairing and stacking of bases indicated by Raman profiles from ssRNA viruses are not evident for the φX174 chromosome. Overall, the proteins of the virion contain extensive β-sheet and irregular secondary structures. Fourier deconvolution of the Raman amide I band provides an estimate of the percentage of total β-sheet structure (~60%) in all proteins of the virion. The amide III region of the spectrum confirms that β-sheet and irregular domains are the predominant protein secondary structures. Samples of φX174 concentrated for Raman spectroscopy by either ultracentrifugation or ultrafiltration exhibit nearly identical Raman spectra, indicating that either method can be employed to prepare intact virus without significant loss of DNA or protein components.

The bacteriophage φX174 contains a single-stranded and covalently closed DNA molecule packaged in an isometric capsid. The complete nucleotide sequence of φX DNA and the amino acid sequences of the proteins which it encodes have been determined (Sanger et al., 1978). However, in contrast to the morphologically related single-stranded RNA (ssRNA) plant viruses which have been extensively studied by X-ray diffraction methods, relatively little is known of the secondary structures and tertiary interactions of φX174 components.

The φX174 genome, of molecular weight 1.7×10^6 and 5386 bases, constitutes 25.5% of the virion mass and encodes 10 genes which specify at least 11 gene products (Sanger et al., 1977). Only four of these, viz., products of genes F (M_r 48 400), G (M_r 19 050), H (M_r 34 400), and J (M_r 4200), are present in extracellular virus. The molar ratios of 60F:60G:12H suggest that a morphological subunit composed of proteins F, G, and H is located at the 12 vertices of an icosahedral capsid (Burgess, 1969; Siden & Hayashi, 1974). Proteins G and H contain domains which project as spikes from the capsid surface (Edgell et al., 1969) and function in recognition of the host receptor apparatus (Incardona, 1978). The product of gene J, a small basic polypeptide, is required for DNA packaging which suggests an internal location in the infectious virus (Aoyama et al., 1981). Although models have been advanced for the morphogenesis of φX174 (Hayashi, 1978), the detailed architecture of the virion and its stable precursor particles, as well as the nature of protein–protein and protein–DNA interactions which may be present in these nucleoprotein assemblies, is not known.

Laser Raman spectroscopy can be profitably exploited to determine details of protein and nucleic acid structures and their mutual interactions in viruses and gene regulatory complexes [reviewed in Thomas (1986)]. Recent applications have used data obtained from model nucleic acid crystals, fibers, and solutions to reveal preferred backbone and nucleoside conformations in DNA and RNA of plant and bacterial viruses (Thomas et al., 1983; Verduin et al., 1984; Prescott et al., 1985). In this paper, we report Raman spectra of φX174 and assign many of the individual Raman lines in the spectra to specific DNA and protein subgroups. The results are interpreted to indicate conformational features of φX DNA packaged in the viral capsid and average secondary structures of the viral proteins. The φX DNA structure, as revealed by Raman spectroscopy, is compared and contrasted with results obtained on other isometric viruses.

A structural interpretation of the virus Raman spectrum depends upon successful resolution of a number of structurally informative spectral bands (or "Raman lines") which originate from vibrations of different subgroups of the macromolecules. In complex macromolecular assemblies, including φX174, instrumental resolution of all structurally informative lines of the protein and DNA residues is not possible, in part because of inherently broad natural bandwidths and in part because of extensive overlap of Raman lines of the DNA and protein with one another. However, resolution can be enhanced by mathematical methods, such as Fourier deconvolution, which is well suited to virus applications (Thomas & Agard, 1984;

† Paper 19 in the series Studies of Virus Structure by Laser Raman Spectroscopy. Supported by Grants AI11855 from the National Institutes of Health (G.J.T.) and DMB-8416573 from the National Science Foundation (N.L.I.).
* Correspondence should be addressed to this author.

0006-2960/87/0426-1532$01.50/0

Thomas, 1985). We employ deconvolution here to enhance resolution of multiple components of certain complex band shapes which comprise structure-sensitive lines, referred to as Raman conformation markers. This facilitates semiquantitative estimation of the proportions of different nucleoside and backbone conformers present in DNA and of different types of secondary structure present in the viral protein.

Experimental Procedures

Preparation and Purification of ϕX174. The lysis-defective mutant E*am*3 was grown in *Escherichia coli* C at 37 °C in Cas media as previously described (Incardona et al., 1985). The infected cells were harvested by centrifugation, 150-min postinfection, and lysed with ethylenediaminetetraacetic acid (EDTA)–lysozyme. After sonication and removal of insoluble debris by centrifugation at 27000*g* for 15 min, the crude lysate was centrifuged for 40 min at 50 000 rpm and 5 °C through a 5–25% (w/w) sucrose gradient containing 1.0 M NaCl and 0.0125 M EDTA–borate buffer. Centrifugations were carried out in the VTi50 rotor of a Beckman L5-50 ultracentrifuge. The virus-containing fractions were located by absorbance readings at 260 nm on diluted aliquots. The remaining undiluted fractions were pooled and diluted 10-fold with borate buffer before sufficiently dry, optical-grade CsCl was added to give a final density of 1.36 g/mL. The isopycnic centrifugation was run at 5 °C in the VTi50 rotor at 50 000 rpm, and the virus-containing fractions were immediately dialyzed against borate buffer and subsequently stored at 5 °C until Raman spectra were recorded. Collection of Raman data on these samples was completed within 2 weeks following the final dialysis procedure.

The purified solutions of virus were divided into two equal fractions, each of which was concentrated approximately 100-fold for Raman spectroscopy by a unique protocol. One fraction was pelleted by standard ultracentrifugation, at 100000*g* in the fixed-angle rotor of a Beckman Airfuge, similar to the procedure used previously for ssRNA plant viruses TYMV, BDMV, and CCMV (Verduin et al., 1984; Prescott et al., 1985). The second fraction was neither pelleted by ultracentrifugation nor subjected to other centrifugation procedures, but was concentrated by sieving in a Centricon ultrafiltration molecular sieve (10 000 daltons). The virus concentrates obtained by ultracentrifugation and ultrafiltration methods were subsequently treated identically for loading into Raman cells and spectroscopic analyses, as outlined below.

The viruses of both fractions were examined by denaturing gel electrophoresis, before Raman spectra were recorded. Electrophoresis was carried out on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gels (10% T) in a Hoeffer SE400 slab gel unit at constant voltages of 120 V through the stacking gel and 200 V through the running gel.

Raman Spectroscopy. Sample handling procedures for Raman spectroscopy of viruses generally differ from those employed for simpler molecules and are described in detail elsewhere (Thomas, 1986). Instrumentation used in this work has also been described (Li et al., 1981).

Solutions of ϕX174 containing approximately 80–120 μg of virus/μL of buffer [150 mM NaCl + 20 mM tris(hydroxymethyl)aminomethane (Tris), pH 7.85] were sealed in glass capillary cells (Kimax 34507) maintained at 4 °C before spectroscopic examination and thermostated at 20 °C during the recording of Raman spectra (Thomas & Barylski, 1970). The spectra were excited with 300 mW of 514.5-nm radiation from a Spectra Physics 171-18 argon laser and were recorded by using a spectral slit width of 8 cm^{-1} on a Spex Ramalog V/VI spectrometer under the control of a North Star Horizon II Model HRZ-2-64K-Q microcomputer. Photon counts were read at intervals of 1 cm^{-1} with an integration time of 1 s. The Raman spectra illustrated below are the averages of numerous repetitive scans, usually 8 for the 300–1800 cm^{-1} region and up to 30 for the narrower regions illustrated. The Raman data have been corrected for scattering by the solvent and buffer and for the gently sloping background characteristic of aqueous solution spectra of viruses.

Signal averaging, base-line correction, spectral subtraction, and related routines were preformed on-line with the data acquisition by the North Star computer (Li et al., 1981). Deconvolution procedures were carried out on an IBM-PC/XT microcomputer with an installed numeric data processor (Intel 8087) using software developed in our laboratory and described elsewhere (Thomas & Agard, 1984). Further information on computer programs is available on request (G.J.T.).

Results and Conclusions

Qualitative Features of Raman Spectra and Band Assignments. Since purification of ϕX174 by differential centrifugation in angle-head rotors frequently yields preparations with low specific infectivity (P. Stow and N. L. Incardona, unpublished results), we examined the alternative method of ultrafiltration to concentrate virus samples for Raman spectroscopy. One possible reason for virus inactivation by centrifugation is the high shear forces at the tube walls, which are maximal for the nearly vertical orientation of the Airfuge rotor. These forces could selectively dislodge spike proteins G and H from the nucleocapsid, rendering the particles incapable of binding and initiating DNA injection (Incardona, 1978). In this section, we compare Raman data obtained from viruses concentrated by both centrifugation and filtration methods.

Figure 1 shows Raman spectra in the region 300–1800 cm^{-1} of H_2O solutions of ϕX174 concentrated for Raman spectroscopy by ultracentrifugation (Figure 1a) and ultrafiltration (Figure 1b) methods, and their difference spectrum (Figure 1c). The difference spectrum was obtained with Figure 1a as minuend and Figure 1b as subtrahend after normalization of the Raman intensities of each to the same intensity standard, viz., the integrated intensity of the protein band near 1448 cm^{-1} which arises from all protein methylene groups (CH_2 scissoring mode) and which has been shown to be an acceptable intensity standard in Raman spectra of other ssDNA and ssRNA viruses (Thomas et al., 1983; Verduin et al., 1984).

The virtual equivalence of the protein and DNA Raman intensities in Figure 1a,b which yield a nearly null difference spectrum over the 500–1500 cm^{-1} interval of Figure 1c, indicates that the molecular compositions of ultracentrifuged and ultrafiltered ϕX174 particles are essentially the same. If ultracentrifugation caused shearing of the protein spikes which protrude from the ϕX174 capsid surface, we would expect Raman lines for a number of the protein groups in Figure 1a to be diminished in intensity relative to their counterparts in the spectrum of the ultrafiltered sample of Figure 1b. This would be reflected by a substantial number of *negative* bands in the difference spectrum of Figure 1c at the positions of the appropriate protein group vibrations or, equivalently, by a substantial number of *positive* bands at the positions of DNA vibrations. Since this is clearly not the case in Figure 1c, it is evident that significant removal of spike proteins from the ϕX174 capsid does not occur as a result of the ultracentrifugation employed.

The difference spectrum of Figure 1c contains uncompensated broad bands only of H_2O near 450–500 and 1600–1650

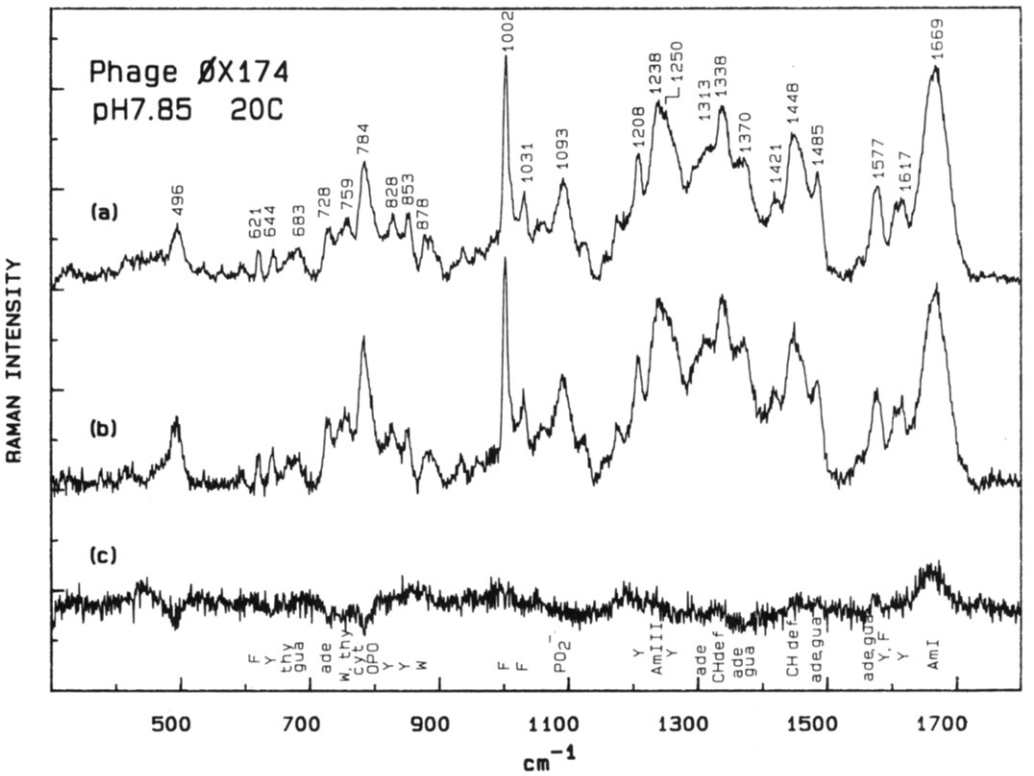

FIGURE 1: Raman spectra in the 300–1800 cm^{-1} region of ϕX174, at 20 °C in 150 mM NaCl + 20 mM Tris, pH 7.85. Other experimental conditions are described in the text. (a) Sample of ϕX174 concentrated to 120 $\mu g/\mu L$ by ultracentrifugation. (b) Sample of ϕX174 concentrated to 80 $\mu g/\mu L$ by ultrafiltration. (c) Difference spectrum (a – b) obtained after normalization of (a) and (b) to yield the same intensity for the 1450 cm^{-1} band. The prominent bands are labeled to indicate Raman frequencies in cm^{-1} units, and labels along the abscissa indicate primary assignments using standard one-letter symbols for the amino acids (upper case), three-letter symbols for the DNA bases (lower case), and abbreviations for skeletal protein (AmI, AmIII) and DNA (OPO, PO_2^-) assignments.

cm^{-1} due to imperfect subtraction of the solvent background, but no conspicuous protein or DNA difference bands. This indicates that molecular groups of the virus and their configurations are virtually identical for the two methods of sample preparation. If differences exist, they are below the present level of detection (i.e., <5%) or are of such a nature as to cancel fortuitously in the subtraction procedure. We consider the latter possibility unlikely and conclude that the ultracentrifugation of ϕX174 at 100000*g* results in no major damage to the virion.

The above conclusion is supported by polyacrylamide gel electrophoresis (PAGE) of ϕX174 samples on sodium dodecyl sulfate (SDS) denaturing gels. Figure 2 compares the SDS–PAGE patterns of ϕX174 samples concentrated by either ultrafiltration or ultracentrifugation but otherwise handled identically for Raman spectroscopic examination. The electrophoresis results indicate that the same proteins are present in approximately the same relative amounts for ultrafiltered (lane 2) and ultracentrifuged (lane 3) samples, although the absolute virus concentrations differ in the two cases. The viral proteins are identified as products of viral genes F, H, G, and J, as labeled in Figure 2. (Relatively small amounts of two contaminating proteins, labeled X and Y, are also present in both samples.)

Prominent Raman lines of ϕX174 in Figure 1 are labeled with the frequency shift in cm^{-1} units, accurate to ± 2 cm^{-1} for sharp lines and slightly more for broad lines or shoulders. Assignments to specific DNA or protein groups, as indicated along the abscissa, are based on previous model compound studies (Lord & Thomas, 1967; Prescott et al., 1984; Lord & Yu, 1970; Thomas et al., 1983). Further details of the assignments to virus components have been given elsewhere (Thomas, 1986). In the following sections, we discuss the

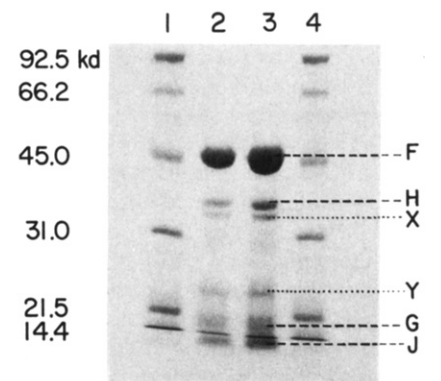

FIGURE 2: SDS–PAGE patterns of ϕX174 samples prepared by the two different methods which yielded the spectra of Figure 1a,b. Lanes 1 and 4 contain the molecular weight standards indicated. Lanes 2 and 3 contain samples originally concentrated by ultrafiltration (80 $\mu g/\mu L$) and ultracentrifugation (120 $\mu g/\mu L$), respectively, and diluted 1000-fold before transfer to the stacking gel. Bands marked F, H, G, and J correspond to the respective viral gene products, which are seen to be present in the same relative proportions even though the absolute concentrations of ϕX174 differ in lanes 2 and 3. The bands labeled X and Y correspond to as yet unidentified proteins (possibly contaminants), with approximate molecular weights of 30 000 and 25 000, respectively.

salient spectral features primarily in relation to the conformation of encapsidated ϕX DNA. Although the spectrum is rich in protein bands, detailed interpretation of them is hindered by the variety of proteins present in the nucleocapsid.

Interpretation of Raman Scattering from Encapsidated ϕX DNA. The ssDNA of ϕX174 contributes numerous Raman lines to the virus spectrum in the intervals 650–840, 1050–1100, 1290–1400, and 1480–1580 cm^{-1} (Figure 1). Those of the first interval are the best resolved from lines of

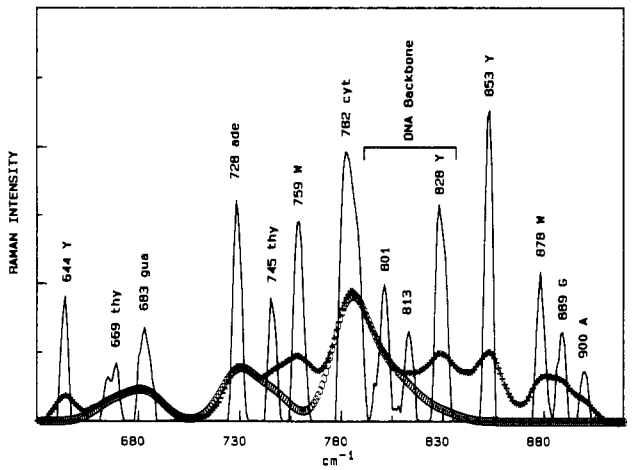

FIGURE 3: Raman spectrum of ϕX174 in the region of conformation-sensitive bands of ϕX DNA. Experimental data (+) after smoothing with a nine-point moving average procedure. Deconvolved data (—) after 50 cycles of iterative deconvolution with a Gaussian function of 20 cm^{-1} half-width (Thomas & Agard, 1984). Also included is a hypothetical spectrum of encapsidated ϕX DNA (○), computed by convolution of the deconvolved DNA bands with the original deblurring function, as described in the text.

viral protein (Thomas et al., 1983), as well as the most informative of DNA secondary structures (Erfurth et al., 1975; Benevides & Thomas, 1983; Prescott et al., 1984). In Figure 3 we show the deconvolution of the smoothed spectral profile of this region. Here we have used the deconvolution procedure described by Thomas and Agard (1984) and applied previously to filamentous viruses (Thomas, 1985). Deconvolution enhances the resolution of all bands in the region and achieves partial separation of protein and DNA components. All of the bands, with the exception of those of tyrosine at 644, 828, and 853 cm^{-1} and that of tryptophan at 878 cm^{-1}, are due in whole or in part to DNA nucleosides or the phosphodiester backbone. Labels of the deconvolved peaks indicate the appropriate assignments.

An important segment of the spectrum of Figure 3 for DNA conformation analysis is the 800–840 cm^{-1} interval (Erfurth et al., 1972, 1975; Thomas & Hartman, 1973; Prescott et al., 1984). Unfortunately, this interval is partly obscured by the lower frequency member of the well-known Fermi doublet of the tyrosine ring (Siamwiza et al., 1975), observed here at 828 cm^{-1} and due to all tyrosines of the viral proteins. A hidden DNA Raman line near 825–835 cm^{-1}, if present, would be indicative of the B form of ϕX DNA and would suggest the formation in the single-stranded ϕX genome of helical "hairpin" segments with the B-backbone geometry. Similarly, a DNA Raman line ca. 807 ± 3 cm^{-1} would indicate the A-form backbone for such helical hairpins, analogous to the secondary structure formed by ssRNA molecules and observed in RNA plant viruses and RNA phages (Thomas, 1986). Conversely, the absence of both the 807 and 830 cm^{-1} DNA markers would constitute evidence for the absence of helical hairpinlike secondary structure. DNA molecules which lack such secondary structure and display no Raman markers ca. 807 or 830 cm^{-1} include heat-denatured DNA and certain poorly ordered polydeoxyribonucleotides, such as poly(dT). Thus, for calf thymus DNA at 90 °C and for poly(dT) at lower temperatures, a weak and diffuse background of Raman intensity is observed in lieu of one or more discrete conformation markers in the 800–850 cm^{-1} interval (Erfurth & Peticolas, 1975; G. J. Thomas, Jr., and J. M. Benevides, unpublished results).

In order to better assess the ϕX DNA backbone conformation from the present data, and specifically to determine whether Raman intensity from DNA underlies the intefering tyrosine line at 828 cm^{-1}, we have compensated for the tyrosine contribution as follows. The intensity of the tyrosine 644 cm^{-1} line (I_{644}) is *not* conformation sensitive (Prescott et al., 1983). Further, the sum of intensities of the components of the tyrosine doublet (i.e., $I_{828} + I_{853}$) is nearly independent of conformation even though the intensities of the individual Fermi components are not (Siamwiza et al., 1975). Examination of Raman spectra of a large number of proteins of both viral and nonviral origins indicates that the quotient $R = I_{644}/(I_{828} + I_{853})$ invariably falls within the narrow range 0.25 ± 0.05 (G. J. Thomas, Jr., unpublished results). [See, for example, spectra published in Verduin et al. (1984), Prescott et al. (1985), Thomas (1986), and references cited therein. Values of R near 0.25 are especially well conserved among several mutants of the λ repressor N-terminal fragment (residues 1–102), even though the mutants differ greatly in the number and hydrogen-bonding environments of their tyrosines (Weiss et al., 1986).] Hence, using a mean value for R of 0.25, together with the observed values of I_{644} and I_{853} from Figure 3, we calculate the expected tyrosine contribution at 828 cm^{-1} to be $I_{828} = 4.0I_{644} - I_{852} = 0.86I_0$, where I_0 is the intensity actually observed at 828 cm^{-1} (arbitrary units, Figure 3). In other words, 86% of the Raman intensity at 828 cm^{-1} in the observed spectrum of Figure 3 can be accounted for by tyrosine alone. Consequently, no more than 14% of the observed Raman intensity at this frequency can be reasonably assigned to the ϕX DNA backbone. This provides a conservative upper limit to the ϕX DNA contribution at ca. 830 cm^{-1} arising from phosphodiester conformers of the B-DNA family.

With protein contributions to the deconvolved spectrum of Figure 3 thus in hand, the protein bands may be subtracted to yield the set of deconvolved DNA bands. The latter in turn may be convoluted with the original deblurring function to generate a hypothetical Raman spectrum corresponding to that of the encapsidated ϕX genome. This hypothetical "encapsidated DNA" spectrum, obtained by convolution, is shown in Figure 3 (open circles) and is further discussed below.

The above computed Raman intensity at 830 cm^{-1} for ϕX DNA is weaker by a factor of 6 or 7 than that observed for double-stranded DNA (dsDNA) and other B-form DNA models (Thomas et al., 1986), indicating that at most 14–17% of the phosphodiester linkages of encapsidated ϕX DNA maintain a geometry similar to that of B DNA. The basis for this intensity comparison is normalization to the 1093 cm^{-1} band of the phosphodioxy anion (Prescott et al., 1984). Figure 3 also indicates that the substantial Raman intensity contained as a diffuse wing on the high-frequency side of the very intense 782 cm^{-1} band yields peaks at 801 and 813 cm^{-1} upon deconvolution. The significance of these deconvolved peaks is not clear. They may simply be artifacts of deconvolution, since no clear-cut shoulders are evident from the experimental data. Nevertheless, the parent intensity from which they are derived is real and analogous to that of heat-denatured DNA. Thus, the 85% or so of backbone OPO groups of ϕX DNA not in the B form presumably adopt a broad range of conformational geometries similar to those of heat-denatured DNA, which also gives rise to a diffuse shoulder on the high-frequency side of its 782 cm^{-1} band (Erfurth & Peticolas, 1975).

Finally, Figure 3 shows that the conformation-sensitive ring modes of thymidine, deoxyguanosine, deoxyadenosine, and deoxycytidine are positioned at 669, 683, 728, and 782 cm^{-1}, respectively, which are Raman markers diagnostic of C2′-endo sugar pucker and anti base orientation about the glycosidic

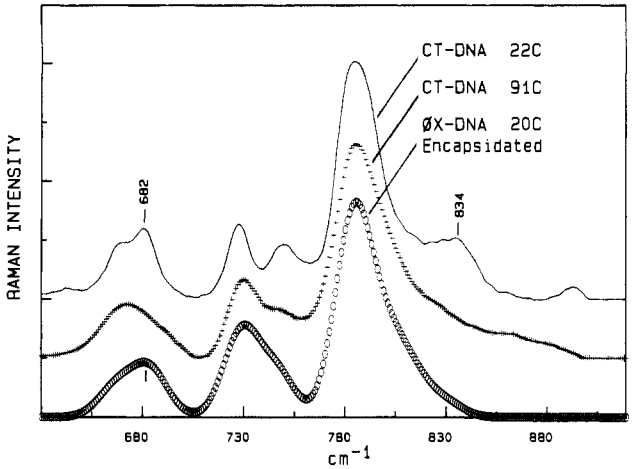

FIGURE 4: Comparison of the Raman spectrum of encapsidated ϕX DNA with Raman spectra of model single- and double-stranded DNAs of similar base composition, for the spectral region 600–900 cm^{-1}. (O) Encapsidated ϕX DNA at 20 °C, obtained as shown in Figure 3. (+) Single-stranded DNA, from calf thymus, denatured at 91 °C. (—) Double-stranded DNA, from calf thymus, at 22 °C.

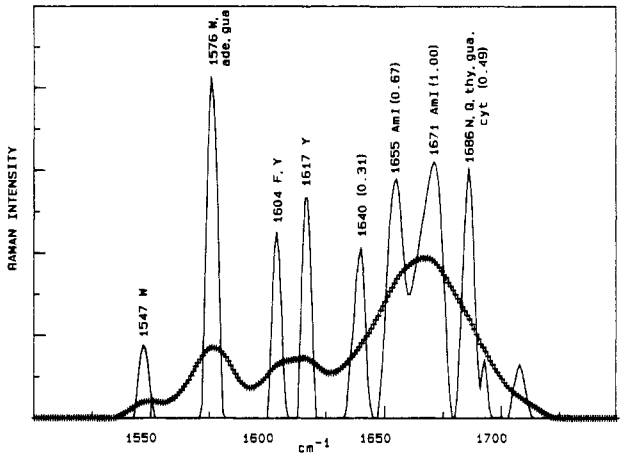

FIGURE 5: Raman spectrum of ϕX174 in the region of the amide I band. Experimental data (+) after smoothing with a nine-point moving average procedure. Deconvolved data (—) after 1000 cycles of iterative deconvolution with a Gaussian function of 25 cm^{-1} half-width (Thomas & Agard, 1984). Labels indicate frequencies of the deconvolved peaks and probable assignments. Integrated intensities are conserved during deconvolution and are indicated in parentheses (arbitrary units) for the major deconvolved peaks of the amide I interval.

bonds of these nucleosides (Thomas et al., 1986). We note also that the additional thymidine marker at 745 cm^{-1} (Figure 3) exhibits the intensity and frequency characteristic of C2′-*endo*/*anti*-thymidine, in contrast to the much greater intensity at 777 cm^{-1} characteristic of C3′-*endo*/*anti*-thymidine (Thomas & Benevides, 1985). Therefore, we find predominantly C2′-endo/anti nucleosides ("B-family" conformers) in ϕX DNA, even though the phosphodiester backbone does not exhibit a typical B-form secondary structure.

In Figure 4, we compare the convolution-generated Raman spectrum (600–900 cm^{-1} region) of the encapsidated ϕX174 DNA molecule with the observed Raman spectra of double-stranded and single-stranded DNA models of similar base composition. This illustration shows both the marked differences between ϕX DNA and dsDNA and the significant similarities between ϕX DNA and heat-denatured DNA.

Interpretation of Raman Scattering from Proteins of ϕX174. The F, G, H, and J proteins together account for 75% of the virion mass and for the most prominent Raman lines of ϕX174. This includes the sharp 1002 cm^{-1} line of phenylalanine residues, which is the single most intense line in the spectrum, as well as the broad band centered near 1669 cm^{-1} which exhibits the greatest integrated intensity and comprises the various backbone amide I modes of all peptide groups. Although DNA bases contain carbonyl groups which should yield collectively a broad band centered near 1670–1680 cm^{-1} (Lord & Thomas, 1967; Prescott et al., 1984), the expected intensity from DNA C=O groups is about one-third that of the strong DNA band ca. 784 cm^{-1} in Figure 1, or about one-sixth that of protein amide I. The DNA bases are therefore only minor contributors to the broad 1669 cm^{-1} band of Figure 1. Thus, the amide I peak at 1669 cm^{-1} and the amide III peaks at 1238 and 1250 cm^{-1} are due primarily to β-sheet and irregular secondary structures (Lord, 1977). Their prominence indicates that *β-sheet and irregular conformations are the prevalent secondary structures of the viral proteins.* The greater intensity of the 1238 cm^{-1} amide III peak indicates further that the percentage of β-sheet structure exceeds that of the irregular domains which generate the 1250 cm^{-1} amide III shoulder (Lord, 1977).

Amide I and III contributions from α-helical domains of the viral proteins are expected near 1645–1655 and 1275–1300 cm^{-1}, respectively (Lord, 1977). However, interpretation of the ϕX174 spectrum in these regions is complicated by two factors. First, the large bandwidth of the 1669 cm^{-1} line would prevent the resolution by instrumental means of a shoulder near 1650 cm^{-1}, even if present. Second, interference from Raman lines of DNA near 1250–1300 cm^{-1} (Prescott et al., 1984) precludes straightforward assignment of putative α-helix amide III bands in this region. Nevertheless, the position of the amide I peak at 1669 cm^{-1} without a prominent shoulder near 1650 cm^{-1} and the presence of a clear-cut minimum at 1275 cm^{-1} together show that the proteins of ϕX174 are overall much lower in total α-helix than in other secondary structure types. This interpretation applies to the net secondary structure of all proteins of the virion, including as many as 60 copies each of F, G, and J, and 12 of H. It does not exclude the possibility that any one of the proteins may contain significant α-helical secondary structure.

We note also that a firm basis exists for the assignment of the weak shoulder near 1265 cm^{-1} to tyrosine and of the more intense shoulders at 1296, 1313, and 1338 cm^{-1} to DNA bases and CH deformations, rather than to protein amide III modes (Lord & Yu, 1970; Lord & Thomas, 1967; Prescott et al., 1984).

In an attempt to obtain a semiquantitative description of the protein secondary structures from the amide I profile, we have employed Fourier deconvolution of the overlapping spectral bands of the 1500–1750 cm^{-1} region, as shown in Figure 5. The assignment of deconvolved peaks at 1547 (W), 1576 (W + gua + ade), 1604 (F + Y), and 1617 cm^{-1} (Y) to aromatic residues is straightforward, but considerable caution must be exercised in interpreting the remaining peaks at 1640, 1655, 1671, and 1686 cm^{-1} as possible amide I bands (Figure 5). Of these, the lowest frequency band at 1640 cm^{-1} (relative intensity I = 0.31) is at the lower limit of the amide I range (Lord, 1977) and is considered to arise primarily from H_2O that cannot be perfectly compensated by the solvent subtraction procedure. [See Verduin et al. (1984) for a further discussion of this problem.] A small contribution from DNA bases is also expected near 1640 cm^{-1} (Prescott et al., 1984). Thus, we expect and assign little or no amide I contribution at 1640 cm^{-1}. Similarly, the band at 1686 cm^{-1} (I = 0.49) is just above the upper cm^{-1} limit of the amide I region and probably arises more from carbonyl group vibrations of glu-

Table I: Predicted Secondary Structures of ϕX174 Proteins[a]

gene product (residues)	CF method				GOR method I				GOR method II			
	% H	% S	% T	% C	% H	% S	% T	% C	% H	% S	% T	% C
F (427)	36.3	33.3	16.9	13.6	17.1	30.5	24.8	27.6	26.0	49.9	12.9	11.2
G (175)	26.9	46.3	18.3	8.6	14.3	45.1	25.2	15.4	24.0	53.7	14.9	7.4
H (328)	52.1	21.0	18.0	8.8	46.7	20.1	12.2	21.0	56.7	29.9	6.4	7.0
J (38)	23.7	10.5	31.6	34.2	7.9	28.9	39.5	23.7	7.9	55.3	23.7	13.1
weighted average[b]	34.7	34.1	18.2	13.0	18.7	33.1	24.5	23.7	27.4	49.2	13.4	10.0

[a] Abbreviations: CF, method of Chou and Fasman (1978); GOR, method of Garnier et al. (1978) with decision constants DCH = 0 and DCS = 0 for GOR-I and DCH = −100 and DCS = −87.5 for GOR-II. H, S, T, and C are the predicted percentages of helix, sheet, turn, and coil, respectively. Computations were carried out as described by Finer-Moore et al. (1984) using microcomputer (IBM-PC) programs which are available on request. [b] Weighted averages are based on 60 copies each of F, G, and J and 12 copies of H per virion.

tamine and asparagine side chains (Thomas et al., 1983), as well as from C=O groups of DNA bases (Prescott et al., 1984), than from protein amide I vibrations. The viral proteins of ϕX174, especially gpH, are in fact quite rich in glutamine and asparagine residues (Sanger et al., 1978). Thus, we attribute most of the deconvolved 1686 cm^{-1} band to other than amide I vibrations.

Accordingly, only the deconvolved peaks at 1655 (relative intensity $I = 0.67$) and 1671 cm^{-1} ($I = 1.00$) are reasonably assigned to protein amide I modes. The latter is at the position expected for β-sheet structures, and the former may contain contributions from both irregular and helical domains. Assuming only these two amide I band components, the deconvolution results thus suggest an upper limit of approximately 60% β-sheet structure (including β-turns) among all viral proteins. A more conservative interpretation would allow for a minor contribution from irregular domains also to the 1671 cm^{-1} component. We have noted earlier that the amide III profile provides additional evidence for a majority of β-sheet structure in ϕX174 viral proteins. We make no effort to further characterize qualitatively or quantitatively the protein secondary structures from the present data. Further analysis must await isolation and purification of individual gene products and partially assembled capsids.

Discussion and Summary

The Raman spectrum of ϕX174 indicates the absence of a substantially ordered backbone of either the B or the A geometry for encapsidated ϕX DNA, which implies that neither extensive regions of stacked bases nor complementary intrastrand base pairing is a major component of the secondary structure of this ssDNA genome. Since substantial base pairing and base stacking, in the form of helical hairpin domains, are expected and observed for other protein-free ssDNA molecules (Nussinov & Pieczenik, 1984a,b), we conclude that the absence of such secondary structure in the DNA of ϕX174 is the consequence of capsid (protein) imposed structural constraints. This result is consistent with previous studies of ϕX174 and strengthens a possible interpretation of recent eclipse kinetic data (Incardona & Muller, 1985).

The eclipse studies revealed that the insertion of 163 bases into the 36-base J–F intercistronic region increased the sensitivity of ϕX174 to urea inactivation at 37 °C but had no effect on eclipse kinetics in vivo. However, deletion of only 27 bases from the same intercistronic region increased the eclipse rate by 50% at all temperatures without altering the inactivation in urea. Together, the Raman and eclipse findings indicate that the J–F intercistronic region is more likely to be involved in protein–DNA interactions than DNA–DNA (intrachain) interactions. Thus, deletion of a small number of bases apparently reduces or eliminates specific protein–DNA interactions which are important for stabilization of the virion, whereas addition of a relatively large number of bases has a lesser impact upon the already stabilized nucleocapsid.

The Raman data of ϕX174 differ greatly from spectra obtained on isometric ssRNA plant and bacterial viruses [reviewed in Thomas (1986)] for which the RNA genomes contain up to 80–95% of nucleosides in ordered conformations (stacked and/or paired, including hairpin helical domains and tertiary interactions). It could be argued that the deficiency of secondary structure in ϕX DNA is derived in part from the greater flexibility of the deoxyribonucleotide backbone vis-à-vis that of RNA. In our view, however, the present finding is more likely the result of specific association between the DNA genome and viral proteins or protein subdomains of the capsid interior wall. The latter explanation would also be more consistent with the demonstrated ability of other single-stranded polydeoxyribonucleotides to form ordered secondary structures even in the absence of a complementary strand.

Recently it was shown, by use of a nucleic acid folding algorithm (FOLD-A), that the minimum energy configuration of the isolated ϕX174 genome is a predominantly base-paired secondary structure comprising mainly helical hairpin domains (Nussinov & Pieczenik, 1984b). In this minimum energy model, the overwhelming majority of backbone phosphodiester groups are expected to assume the helical geometry characteristic of B-form DNA (Nussinov & Pieczenik, 1984a). The present results, therefore, are interpreted as evidence for the existence of constraints within the capsid, most likely imposed by proteins, to prevent formation of a significantly ordered (base-paired) secondary structure in the ϕX174 chromosome. At least two of the four capsid proteins are likely candidates: There is an absolute requirement for J protein in the in vitro packaging of ϕX DNA into infectious virions (Aoyama et al., 1981), and more than one molecule of H protein is injected into the host cell along with the DNA during infection (Jazwinski et al., 1975).

The Raman spectra of ϕX174 indicate extensive secondary structure in capsid proteins collectively but do not provide detailed information on the types of structures in proteins F, G, H, and J, individually. However, since the viral protein sequences are known (Sanger et al., 1978), their sequence-preferred secondary structures may be assessed using predictive methods. In Table I, we summarize predicted secondary structures from methods of Chou and Fasman (1978) (CF) and Garnier et al. (1978) (GOR). In the latter case, two calculations were performed: set I for which the "decision constants" were unbiased (i.e., DCH = DCS = 0), and set II for which the decision constants were biased in favor of ordered structures (DCH = −100, DCS = −87.5). The effect of the latter biasing is to increase the predicted amount of helix in all proteins, mainly at the expense of coil and turn. The algorithms of Garnier et al. and Chou and Fasman yield similar results for gpH, which is predicted to be predominantly α-helical, and for gpG, which is predicted to be predominantly β-sheet. The two methods give quite dissimilar results for gpF

and gpJ, and the CF predictions overall are in better agreement with the present experimental findings. Isolation of capsid subunits in quantities sufficient for Raman spectroscopy is in progress, and results of further examination of the four proteins separately will be reported in a future paper.

ACKNOWLEDGMENTS

We thank Dr. James M. Benevides for discussions and technical assistance.

REFERENCES

Aoyama, A., Hamatake, R. K., & Hayashi, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7285–7289.

Benevides, J. M., & Thomas, G. J., Jr. (1983) *Nucleic Acids Res. 11*, 5747–5761.

Burgess, A. B. (1969) *Proc. Natl. Acad. Sci. U.S.A. 64*, 613–617.

Chou, P. Y., & Fasman, G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Edgell, M. H., Hutchison, C. A., III, & Sinsheimer, R. L. (1969) *J. Mol. Biol. 42*, 547–557.

Erfurth, S. C., & Peticolas, W. L. (1975) *Biopolymers 14*, 247–264.

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

Erfurth, S. C., Bond, P. J., & Peticolas, W. L. (1975) *Biopolymers 14*, 1245–1257.

Finer-Moore, J., Stroud, R. M., Prescott, B., & Thomas, G. J., Jr. (1984) *J. Biomol. Struct. Dyn. 2*, 93–100.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Hayashi, M. (1978) in *The Single Stranded DNA Phages* (Denhardt, D. T., Dressler, D., & Ray, D. S., Eds.) pp 531–547, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Incardona, N. L. (1978) in *The Single Stranded DNA Phages* (Denhardt, D. T., Dressler, D., & Ray, D. S., Eds.) pp 549–555, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Incardona, N. L., & Muller, U. R. (1985) *J. Mol. Biol. 181*, 479–486.

Incardona, N. L., Tuech, J. K., & Murti, G. (1985) *Biochemistry 24*, 6439–6446.

Jazwinski, S. M., Marco, R., & Kornberg, A. (1975) *Virology 66*, 294–305.

Li, Y., Thomas, G. J., Jr., Fuller, M., & King, J. (1981) *Prog. Clin. Biol. Res. 64*, 271–283.

Lord, R. C. (1977) *Appl. Spectrosc. 31*, 187–194.

Lord, R. C., & Thomas, G. J., Jr. (1967) *Spectrochim. Acta, Part A 23A*, 2551–2591.

Lord, R. C., & Yu, N.-T. (1970) *J. Mol. Biol. 50*, 509–524.

Nussinov, R., & Pieczenik, G. (1984a) *J. Theor. Biol. 106*, 245–259.

Nussinov, R., & Pieczenik, G. (1984b) *J. Theor. Biol. 106*, 261–273.

Prescott, B., Steinmetz, W., & Thomas, G. J., Jr. (1984) *Biopolymers 23*, 235–256.

Prescott, B., Sitaraman, K., Argos, P., & Thomas, G. J., Jr. (1985) *Biochemistry 24*, 1226–1231.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III, Slocomb, P. M. & Smith, M. (1977) *Nature (London) 265*, 687–695.

Sanger, F., Coulson, A. R., Freidmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M., & Smith, M. (1978) *J. Mol. Biol. 125*, 225–246.

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

Siden, E. J., & Hayashi, M. (1974) *J. Mol. Biol. 89*, 1–16.

Thomas, G. J., Jr. (1985) *Spectrochim. Acta, Part A 41A*, 217–221.

Thomas, G. J., Jr. (1986) *Adv. Spectrosc. 13*, 233–309.

Thomas, G. J., Jr., & Barylski, J. R. (1970) *Appl. Spectrosc. 24*, 463–464.

Thomas, G. J., Jr., & Hartman, K. A. (1973) *Biochim. Biophys. Acta 312*, 311–322.

Thomas, G. J., Jr., & Agard, D. A. (1984) *Biophys. J. 46*, 763–768.

Thomas, G. J., Jr., & Benevides, J. M. (1985) *Biopolymers 25*, 1101–1105.

Thomas, G. J., Jr., Prescott, B., & Day, L. A. (1983) *J. Mol. Biol. 165*, 321–356.

Thomas, G. J., Jr., Benevides, J. M., & Prescott, B. (1986a) in *Biomolecular Stereodynamics IV* (Sarma, R. H., & Sarma, M. H., Eds.) pp 227–254, Adenine Press, Guilderland, NY.

Thomas, G. J., Jr., Prescott, B., Benevides, J. M., & Weiss, M. A. (1986b) *Biochemistry 25*, 6768–6778.

Verduin, B. J. M., Prescott, B., & Thomas, G. J., Jr. (1984) *Biochemistry 23*, 4301–4308.