C2, an Unusual Filamentous Bacterial Virus: Protein Sequence and Conformation, DNA Size and Conformation, and Nucleotide/Subunit Ratio[†,‡]

Leondios G. Kostrikis,[§] Stephen A. Reisberg,[∥] Hye-Young Kim,[⊥] Sunny Shin,[#] and Loren A. Day*

The Public Health Research Institute, 455 First Avenue, New York, New York 10016

Received August 31, 1994; Revised Manuscript Received December 28, 1994[⊗]

ABSTRACT: *Inovirus* C2 is 1295 nm long and 6.8 nm in diameter, and its mass is 24 million Da. Its genome is a topologically circular, single-stranded DNA molecule of 8100 nucleotides. The DNA is packed in the virion as two antiparallel strands, with a rise per nucleotide in each strand of 3.2 Å; it can be assigned spectroscopic properties like those of base-stacked, right-handed, double-stranded DNA. The stoichiometric ratio (*n/s*) of nucleotides to subunits of the major coat protein is close to 2. The protein subunit contains 52 amino acids, and the DNA sequence of its gene does not encode a signal peptide. The protein conformation in the virion is helical, mostly α-helix with perhaps some 3_{10}-helix. The amino acid sequence of the DNA interaction domain of the subunit is unique among *Inovirus* species. On the basis of its coat protein sequence and available theories of helical symmetry in such structures, C2 appears to be either an unusual member of filamentous virus symmetry class II or the defining member of a new symmetry class.

C2 is a virus of the genus *Inovirus*, the members of which are flexible filaments on the order of 10^3 nm in length, but less than 10 nm in diameter. Each of these viruses contains a topologically circular, single-stranded DNA molecule within a shell of highly α-helical coat protein subunits. The loop of DNA extends from one end of the virion to the other, so that there are two antiparallel strands. X-ray fiber diffraction patterns and mass-per-length measurements have yielded the fundamental helical symmetries of the protein coats of six such viruses, on the basis of which they fall into two symmetry classes (Marvin *et al.*, 1974a,b; Wiseman & Day, 1977; Newman *et al.*, 1977, 1982; Nave *et al.*, 1981; Makowski & Caspar, 1981; Peterson *et al.*, 1982). For each virus in class I (Ff, IKe, and If1), the protein subunit helix has a 5-fold rotational axis with an approximate 2-fold screw, the symmetry being designated $C_5S_{\sim 2}$. In class II viruses (Pf1, Xf, and Pf3), the protein helices have approximately 5.4-fold screw symmetry, nominally $C_1S_{\sim 5.4}$. The subunits in successive levels of pentamers in class I virions, or the subunits in successive helical turns in class II virions, interdigitate to form the cylindrical shells around the DNA, with 10 (class I) or 11 (class II) α-helical segments of different subunits forming the inner wall at each point [see, for example, Makowski *et al* (1980) and Marzec and Day (1988)]. Such viruses are assembled at the host cell membrane as they are extruded into the medium, and various aspects of these processes have been considered recently by several authors (Russel, 1991, 1993; Guy-Caffey *et al.*, 1992; Thiaudiere *et al.*, 1993; Kazmierczak *et al.*, 1994).

The DNA structures within the virions in the genus *Inovirus* show more diversity than the protein coat structures, and strains of the same protein symmetry class can have widely differing DNA structures [for a review, see Day *et al.* (1988)]. Dramatic changes in virion length, and hence changes in DNA structure, can be induced by directed mutagenesis of the class I virus Ff(fd) major coat protein (Hunter *et al.*, 1987). Even more dramatic changes in DNA structure have occurred over the course of evolution, as is evident in comparisons of the class II viruses Xf and Pf1, which have extensive sequence homology. In Pf1 the DNA and protein helices both have pitches of about 16 Å, whereas in Xf the subunit helix also has a pitch of about 16 Å, but the DNA helix has a pitch near 30 Å (Casadevall & Day, 1983; Kostrikis *et al.*, 1994; Liu & Day, 1994; Marzec & Day, 1994).

C2 is an IncC plasmid-specific virus of Gram-negative bacteria (Bradley *et al.*, 1982). It was shown by electron microscopy to have different end structures, a stronger tendency to aggregate, a different mode of pili attachment, and a shorter persistence length in comparison to other filamentous phages (Bradley *et al.*, 1982; Reisberg, 1989). A study of C2 structure was undertaken because the observed macroscopic differences suggested that C2 might differ in important ways from the known viruses of this type. In the current study, the virus was purified, its major coat protein gene (gene VIII) was identified and sequenced, and the virus length, mass-per-length, stoichiometry, genome size, and conformations of its DNA and protein were all determined. The combined set of data indicates a special role for C2 in investigations of the patterns of structural diversity and similarity already evident among species of the genus *Inovirus*.

[†] Financial support was through NIH Grant GM42286.

[‡] From part of the Ph.D. Thesis of L.G.K. in the Department of Chemistry, New York University, New York, NY.

* Author to whom correspondence should be addressed.

[§] Present address: Aaron Diamond AIDS Research Center, 455 First Ave., New York, NY 10016.

[∥] Deceased.

[⊥] Present address: Ewha Womans University, Dahyundong 11-1, Seoul, Korea.

[#] Present address: Massachusetts Institute of Technology, Cambridge, MA 02139.

[⊗] Abstract published in *Advance ACS Abstracts,* March 1, 1995.

0006-2960/95/0434-4077$09.00/0

MATERIALS AND METHODS

Bacterial Growth and Virus Purification. C2 virus and strains of *Salmonella typhimurium* LT2 bearing IncC plasmids that serve as hosts were obtained from D. E. Bradley (Memorial University of Newfoundland, St. John's, Newfoundland, Canada). *S. typhimurium* was grown in BHI medium (200 g of calf brain infusion, 250 g of beef heart infusion, and 2.5 g of disodium phosphate per liter). The bacterial growth was monitored by light-scattering measurements at 600 nm and/or direct counting by microscopy using a Petroff−Hauser chamber. Cells grown to a density of 10^8 bacteria/mL were infected with C2 virus. The multiplicities of infection were about 10−20 PFU/cell. The infected cultures were incubated at 37 °C for about 16 h. The cells were removed by centrifugation, and the supernatant solution was made 1 M in NaCl and 10% in poly(ethylene glycol) and allowed to stand for 2−3 h at room temperature in order for the precipitate to develop (Yamamoto *et al.*, 1970). The virus precipitate was then collected by centrifugation at 16000*g* for 20 min and resuspended in 10 mM sodium phosphate buffer (pH 7.4). Aliquots of 1 mL were layered on 5−20% linear sucrose gradients in the same buffer and centrifuged in a Beckman SW 41 rotor at 40000*g* for 1.5−3 h. The gradients were drained with monitoring by ultraviolet absorption at 270 nm. The virus fractions were dialyzed against 10 mM sodium phosphate buffer (pH 7.4), and their titer and UV absorbance were determined. C2 aggregation in CsCl and KBr precluded the use of these salts for buoyant density gradient purifications, so the virus characterized here was subjected to additional purification based on sedimentation velocity, including two rounds of pelleting differential sedimentation and a repeat sucrose gradient fractionation. Yields were approximately 2 mg of purified virus per liter of infected culture. Increased yields of virus were obtained from *Escherichia coli* JM109 transfected with C2 dsDNA[1] (Kostrikis, 1993). As originally observed by Bradley *et al.* (1982), we found C2 plaques on *S. typhimurium* to be of variable size, sometimes difficult or impossible to see. Incubation of the plates at room temperature produced the best results.

Purification of Viral ssDNA, dsDNA, and Protein. The C2 replicative form dsDNA was isolated from C2-infected *S. typhimurium* according to a protocol for plasmid DNA (Maniatis *et al.*, 1982). After an overnight incubation at 37 °C, the infected cells were harvested by sedimentation at 10 K for 10 min, chilled on ice, washed twice with cell buffer (10 mM Tris-HCl (pH 8.1) and 70 mM NaCl), and lysed with lysozyme at 150 μg/mL and 2% Triton X-100. Supercoiled C2 dsDNA was then purified by CsCl−EtBr equilibrium density gradient ultracentrifugation (Radloff *et al.*, 1967), followed by ethanol precipitation. The DNA pellet was dissolved in TE (10 mM Tris-HCl (pH 8.0) and 0.1 EDTA) and analyzed by agarose gel electrophoresis. High multiplicities of infection (MOI) gave better yields of supercoiled C2 dsDNA, but the yields were lower than for other filamentous phage systems. Increased yields were obtained from *E. coli* JM109 transfected with C2 dsDNA (Kostrikis, 1993).

C2 ssDNA was isolated from purified C2 virus by repeated phenol extractions, followed by repeated ethanol precipitations, and was redissolved and stored in 10 mM Tris-HCl and 0.1 mM EDTA (pH 8.0). The size of the C2 DNA was determined by alkaline agarose gel electrophoresis in running buffers containing 30 mM NaOH and 1 mM EDTA at gel concentrations of 0.6%, 0.8%, and 1.0% agarose, as described by Shin and Day (1995). Gels were run at 5, 4, or 1 V cm^{-1} for 3, 6, or ~15 h, respectively, and neutralized in 1 M Tris-HCl, 1.5 M NaCl, and 0.5 μg/mL EtBr (pH 7.6), and the DNA bands were visualized and photographed under UV light. DNA size standards were the circular ssDNA genomes of Pf3 (5833 bases), M13mp18 (7250 bases), M13mp18 with a 593 base insert (7843 bases), and MC70 (9601 bases), which is fd (6408 bases) with a 3193 base insert. C2 protein was isolated from the phenolic phases during ssDNA isolation according to a standard procedure used for TMV, fd, and Pf1 proteins (Anderer, 1959; Knippers & Hoffmann-Berling, 1966; Kostrikis *et al.*, 1994).

SDS/Urea−Polyacrylamide Gel Electrophoresis. The protein composition of the virus was examined using SDS/urea−polyacrylamide gel electrophoresis according to a protocol from Bethesda Research Laboratories for proteins in the 3−40 kDa range. A resolving gel of 15% acrylamide (0.4% bisacrylamide) and a stacking gel of 3.5% acrylamide was used. Both gels contained 0.1% SDS, 6 M urea, and 0.1 M sodium phosphate (pH 7.2). Polymerization was catalyzed by 0.025% ammonium persulfate and *N,N,N′,N′*-tetramethylenediamine (TEMED). The elution buffer was 0.1 M sodium phosphate (pH 7.2) and 0.1% SDS. Samples of intact C2 virus were loaded onto the gels in 10 mM sodium phosphate, 1% SDS, 0.01% bromophenol blue, 1% β-mercaptoethanol, and 7 M urea (pH 7.2). The proteins were fixed/stained in 50% methanol, 10% acetic acid, and 0.25% Coomassie Blue (R250) for 1 h at 60 °C and destained in 50% methanol and 10% acetic acid for several hours at room temperature. Molecular weight standards were prepared protein mixtures (Bethesda Research Laboratories).

Determination of the Primary Structure of the C2 Major Coat Protein. The amino acid sequence of the C2 major coat protein (gene VIII protein) was determined through protein sequencing of the first few amino acids from the N-terminus of the protein (MGPTAPTD) combined with DNA sequencing of its gene. The protein sequencing was performed on a sample of about 20 μg of purified C2 virus. A PI 2090E integrated microsequencing system was used. DNA sequencing was carried out by the chain termination method (Sanger *et al.*, 1980), with sequenase as the enzyme (United States Biochemicals). Band compressions were reduced through the use of dGTP analogs, 7-deaza-dGTP and dITP, in the place of dGTP in the reaction mixtures, and the gel temperatures were maintained at 50 ± 3 °C. The gene for the coat protein (gene VIII) was identified by sequencing on the plus strand DNA isolated from the virus, with the synthetic mixed oligonucleotide d(GGNGCNGTNGGNTCCAT) as primer, synthesized so that A, G, C, and T occur with equal probability at positions labeled N. This 17-mer is the reverse complement (minus one nucleotide) of the reverse translation of the first six amino acids. From the sequence obtained with this first primer, a second primer was synthesized and used to sequence across the gene on the minus strand of the RF DNA. DNA sequences from this primer were then used to synthesize a third primer to

[1] Abbreviations: dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; EtBr, ethidium bromide; EDTA, ethylenediaminetetraacetic acid; TMV, tobacco mosaic virus; SDS, sodium dodecyl sulfate; PFU, plaque-forming units.

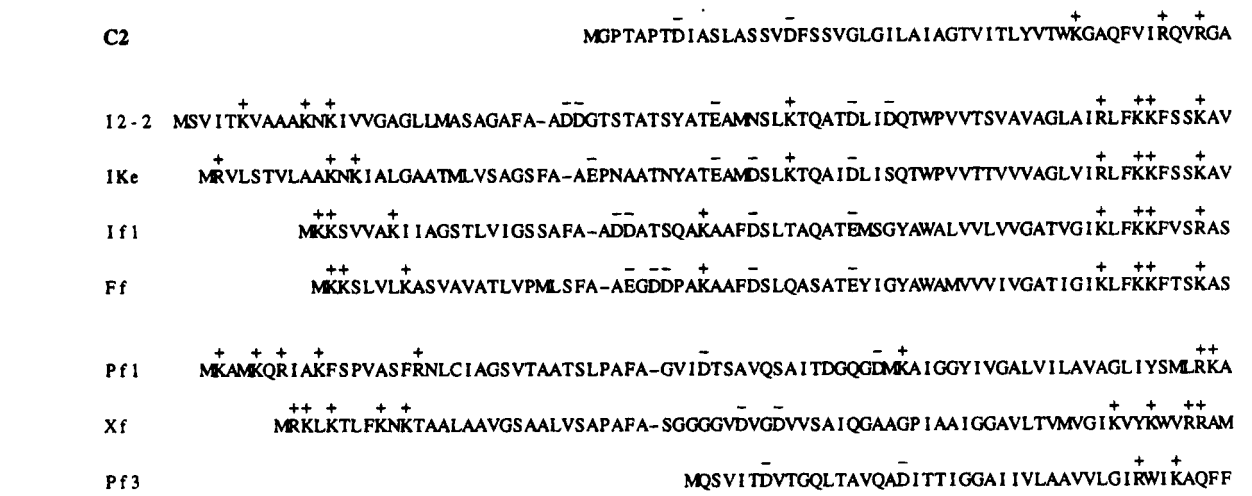

FIGURE 1: Primary structure of C2 gene VIII protein compared to those of other viruses. The sequences are aligned to allow easier comparison of their DNA interaction domains. Basic and acidic residues are indicated by + and −. The signal peptide of each protein, if present, is separated from the mature protein by (−). The C2 sequence was translated from the 156 base open reading frame of its gene, which is, with the start codon in bold and the preceding and following stop codons in italics, *TAA*AGGTGTTTTTTT**ATG**GGACCTACT-GCTCCTACTGATATTGCAAGTTTGGCCTCTTCTGTCGATTTCTCCTCTGTGGGTCTCGGCATTCTGGCTATCG-CTGGCACCGTGATCACTCTGTATGTCACCTGGAAAGGTGCTCAGTTCGTGATTCGCCAGGTGCGCGGCGCT*TAA*. Literature references are as follows: I2-2 (Stassen *et al.*, 1992), IKe (Peeters *et al.*, 1985), If1 (Carne *et al.*, 1991), Ff [fd,f1,M13,ZJ/2; fd is shown (Snell & Offord, 1972; Beck *et al.*, 1978; Van Wezenbeek *et al.*, 1980; Hill & Petersen, 1982)], Pf1 (Hill *et al.*, 1991), Xf [DNA sequence data of H.-Y. Kim, C. K. Bagdassarian, and J. H. Sun, together with the protein sequence data of Frangione *et al.* (1978)]. Pf3 (Putterman, 1983; Luiten *et al.*, 1983; Putterman *et al.*, 1984). Among the Ff species, the signal sequences of f1, fd, and M13 are identical, M13 mature coat has N instead of D at position 12, and ZJ/2 mature coat has A instead of G at position 34.

sequence across the entire gene on the plus strand. The complete gene was sequenced twice in each direction.

Virus Dimensions and Mass from STEM. Virus mass and length determinations were from analyses of micrographs obtained with the scanning transmission electron microscope (STEM) at Brookhaven National Laboratory, with the aid of methods developed for sparsely sampled, curved filamentous particles by S. A. Reisberg (Reisberg, 1989) and J. S. Wall (Wall & Hainfeld, 1986). Micrographs were recorded on unstained virus samples. The mass-per-length of C2 virus was calculated on the basis of tobacco mosaic virus (TMV) as standard. C2 total mass was obtained as length multiplied by mass-per-length.

Ultraviolet Absorption and Circular Dichroism Spectroscopy. Digital UV absorbance spectra with values every 0.1 nm were recorded on a Cary 219 spectrophotometer interfaced to a 386 SX20 PC and then transferred to a Unix-based computer system for analysis and display. CD spectra were recorded on a Cary 60 spectropolarimeter with a CD attachment. The instrument was calibrated with camphorsulfonic acid-d_{10} (Chen & Yang, 1977), and the analog chart data were digitized by means of a digital scanner with software based on image processing techniques written by D. J. Liu [see Kostrikis *et al.* (1994)]. Ag(I) DNA probing experiments were performed according to the method described by Casadevall and Day (1982, 1983). The concentrations of C2 virus were determined by absorbance measurements with an extinction coefficient of 3.51 ± 0.23 mL mg^{-1} at 265 nm, before light-scattering correction (see the following).

RESULTS

Primary Structure of the Major Coat Protein (Gene VIII Protein). The amino acid sequence of the C2 major coat protein (also called the gene VIII protein in accord with Ff phage genome nomenclature) is shown in Figure 1 with the major coat proteins of seven other *Inovirus* species. The C2 major coat protein is composed of 52 amino acids and has a molecular weight of 5191. The first nine amino acids of the protein were obtained by direct protein sequencing. The C2 protein has an N-terminal domain with only two negatively charged residues (D_8 and D_{17}), a hydrophobic central domain of 22 residues (F_{18}−W_{39}), and a C-terminus having three basic residues (K_{40}, R_{47}, and R_{50}), two of which are followed by glycine (G_{41} and G_{51}). None of the other proteins in Figure 1 has a basic residue followed by glycine. The C2 amino acid sequence shows the same overall structure as the major coat proteins of other viruses, several of which (Ff, IKe, Pf3, and Pf1) have been proven to be membrane proteins prior to their assembly into virions (Smilowitz, 1974; Kuhn *et al.*, 1987, 1990; Greenwood & Perham, 1989; Rohrer & Kuhn, 1990; Shen *et al.*, 1991).

C2 Coat Protein Has No Signal Peptide. Protein sequencing showed the first residue to be methionine. This indicated that the mature major coat protein might be synthesized without a signal peptide. This was confirmed in the DNA sequence of the gene given in the legend to Figure 1, which shows a TAA stop codon only 15 residues upstream from the ATG start codon of the open reading frame for the 52 amino acid protein. C2 provides the second instance of a major coat protein of an *Inovirus* synthesized without a signal (leader) peptide, the first having been Pf3 (Putterman, 1983; Luiten *et al.*, 1983; Putterman *et al.*, 1984; Kuhn *et al.*, 1990).

Genome Size and Virion Major Protein Component. The size of the single-stranded DNA of C2 virus is 8100 ± 200 nucleotides and its topology is circular, according to comigration studies on alkaline agarose gels, with four ssDNA molecules of known sizes and of both linear and circular topology (Table 1). Purified DNA from C2 virus showed two bands in the alkaline gels, the relative amounts and mobilities of which behaved like the linear and circular forms of a single molecular weight species (Shin & Day, 1995).

Table 1: Sizing of C2 Single-Stranded DNA

DNA standards in gel	gel concn (%)	apparent size[a]
M13mp18, M13mp18+593, Pf3	0.8	7950, 8200
M13mp18, M13mp18+593, Pf3	0.8	8200
fd, M13mp18, MC70, Pf3	1.0	8000
M13mp18, MC70, Pf3	0.6	8100, 8300
M13mp18, MC70, Pf3	0.6	7900, 8200
average size		8100 ± 200

[a] The first line has values measured for linear and circular forms resolved on a single gel, the next two lines give average values measured for circular forms, and the last two lines give values for two different lanes under conditions that did not resolve linear and circular forms [see Shin and Day (1995)].

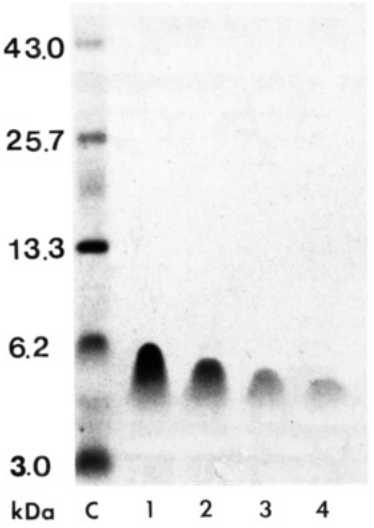

FIGURE 2: Polyacrylamide gel electrophoresis of disrupted C2 virus under gel conditions for low molecular weight proteins in the presence of SDS and urea. In lane C are molecular weight controls. The amounts of C2 virus loaded in lanes 1−4 were 8.8, 4.4, 3.0, and 1.5 μg, respectively.

The major coat protein of C2, known to have a molecular weight of 5191, has an apparent molecular weight of 5000 ± 500 on the polyacrylamide gels for low molecular weight proteins (Figure 2). Overloading of the gels results in aggregates of the major coat protein with higher apparent molecular weights, similar to effects with the fd and Pf3 coat proteins (Putterman *et al.*, 1984). Special efforts to identify minor proteins that are present at the ends of other such viruses (Simons *et al.*, 1981) were not carried out for C2.

Virion Dimensions, Mass, and Stoichiometry from STEM. Scanning transmission electron micrographs of unstained samples (Figure 3) showed C2 virions to be 1295 ± 10 nm long, about 6.8 nm in diameter, and of a uniform mass-per-length of 18 500 ± 250 Da/nm. The value for the length is an average from measurements on images of 46 particles, 32 from a sample of C2 produced by *S. typhimurium* and 14 from a sample produced by *E. coli*. The length is in accord with the results of Bradley *et al.* (1982), who found an average of 1256 ± 84 nm from measurements on 58 images. The mass-per-length was obtained from measurements on 169 viral segments, each about 30 nm long, with a total contour length of about 5000 nm. TMV was the internal mass standard (Wall & Hainfeld, 1986). The total mass of the C2 virion is 24.0 ± 0.5 million Da.

According to the STEM measurements, there are 4130 ± 150 coat protein subunits. This is calculated from the total virion mass, the DNA mass (8100 ± 200 nucleotides multiplied by 310 Da/nucleotide), and the protein subunit mass (5191 Da). From these numbers, we find $n/s = 1.96 \pm 0.09$. The uncertainty is the statistical fluctuation in the data; there could be a systematic error resulting from different relative amounts of retained moisture in the C2 and TMV virions under the conditions of the microscopy. From the length and the number of nucleotides, we calculate an average axial rise per nucleotide of 3.20 ± 0.08 Å. This is simply the length of the virion divided by half the number of nucleotides. From the length and the number of subunits, we calculate an average axial rise per subunit of 3.14 ± 0.12 Å. The STEM data and the parameters derived from them are listed in Table 2.

Stoichiometry from Spectroscopy. A value for *n/s* was obtained by a spectroscopic method that finds the ratio of DNA and protein spectra that best fits the spectrum of disrupted virus in an alkaline urea detergent solution. Small aliquots (25−30 μL) of purified C2 virus (0.5 mg mL^{-1}) were diluted into, and disrupted by, 0.6 mL of (0.2 M NaOH, 1% SDS, and 8 M urea, pH 13.3), and their UV spectra were recorded from 340 down to 250 nm, with data points stored every 0.1 nm (900 points). The spectrum was fitted to reference spectra for the isolated DNA and protein by linear regression. All points were used and all points were plotted without smoothing. Three analyses each were done on two virus preparations; one of the analyses is shown in Figure 4. The fits were not perfect, giving regression correlation

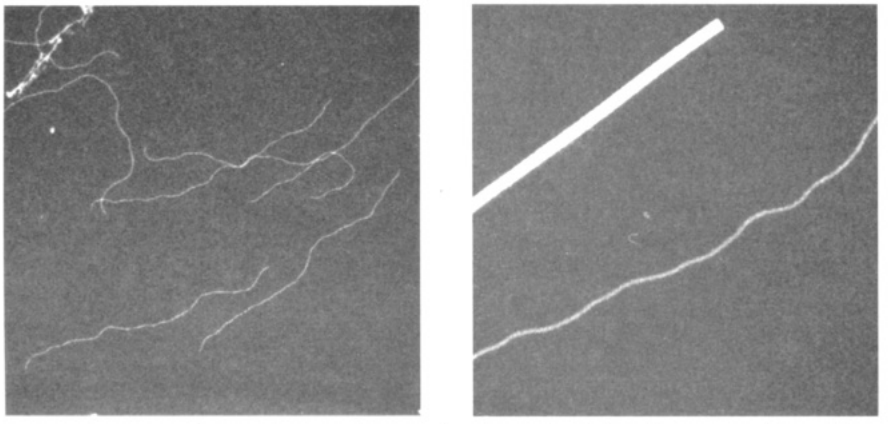

FIGURE 3: Representative scanning transmission electron micrographs of unstained C2 virions on carbon grids used for length (left) and mass-per-length measurements (right). The edge of the micrograph on the left corresponds to 2048 nm, and the edge on the right corresponds to 512 nm. The dense particle in the right micrograph is TMV.

Table 2: Physical Parameters of C2 Virus

length[a] (nm)	1295 ± 10
diameter[a] (nm)	~6.8
mass-per-length[a] (Da nm^{-1})	18500 ± 250
total mass[b] (10^6 Da)	24.0 ± 0.5
total nucleotides[c]	8100 ± 200
total major coat protein subunits[d]	4130 ± 150
n/s from STEM[e]	1.96 ± 0.09
n/s from spectroscopy[f]	2.14 ± 0.13
percent DNA	10.7 ± 0.5
rise per nucleotide[g] (Å)	3.20 ± 0.08
rise per subunit[h] (Å)	3.14 ± 0.12

[a] From STEM. [b] Product of the length and the mass per length. [c] From alkaline agarose gel electrophoresis (Table 1). [d] (Total mass from STEM) − (DNA mass from electrophoresis) divided by subunit mass from sequence. [e] Total nucleotides divided by total subunits (fotnotes *c* and *d*). [f] From regression analyses of absorbance spectra (Figure 4). [g] Length divided by half the number of nucleotides. [h] Length divided by number of subunits (footnote *d*).

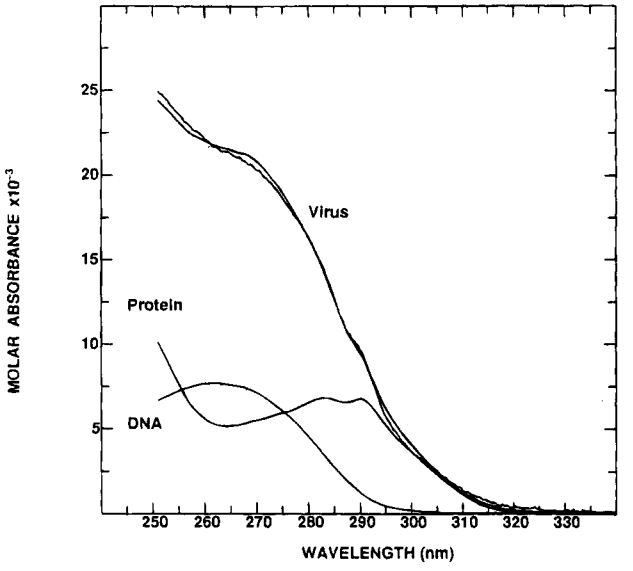

FIGURE 4: Absorbance spectrum of C2 virus disrupted in 0.2 M NaOH, 1% SDS, and 8 M urea (pH 13.3). The noisier of the two spectra labeled virus is the average of three measured spectra of disrupted C2 virus, and the smoother one is the best approximation to it using the DNA and protein reference spectra. The fitting yielded the mole ratio *n/s* through the extinction coefficients ϵ_{DNA} = 7700 ± 200 M^{-1} cm^{-1} at 262 nm and $\epsilon_{protein}$ = 6800 ± 350 M^{-1} cm^{-1} at 284 nm (see text).

coefficients of R^2 = 0.9991 ± 0.0002. Better fits, with R^2 = 0.9999 ± 0.0001, were obtained for the Pf1 virus in a parallel study (Kostrikis *et al.*, 1994).

The fitting ratios yield *n/s* ratios directly because the DNA and protein reference spectra are expressed in extinction coefficients per mole of nucleotide and per mole of subunit, respectively. The uncertainty in *n/s* is primarily due to the uncertainty in the extinction coefficients. We used assigned values for the protein and DNA extinctions at the wavelengths of maximum absorbance because direct absolute measurements required more sample than was available. For the C2 protein subunit, λ_{max} = 284 nm was observed (Figure 4). A value for $\epsilon_{\lambda=284}$ of 6800 M^{-1} cm^{-1} was obtained as the sum of the absolute reference spectra for phenylalanine, tyrosine, and tryptophan under alkaline conditions in the molar ratio 2/1/1 (Mihalyi, 1970). A 5% uncertainty (±350 M^{-1} cm^{-1}) is estimated on the basis of the observations that measured extinction coefficients for the fd and Pf1 proteins by dry weight measurements are within 5% of values calculated from their chromophores [Knippers & Hoffmann-Berling, 1966; Day, 1969; Wiseman & Day, 1977; see also Gill and von Hippel (1989)].

For the average nucleotide in C2 DNA in the alkaline measuring solution, we have assigned ϵ(P) = 7700 ± 400 M^{-1} cm^{-1}. This is the average of alkaline extinction coefficients for two ssDNAs, ϵ(P) = 7830 ± 200 M^{-1} cm^{-1} for Pf1 ssDNA and ϵ(P) = 7500 ± 200 M^{-1} cm^{-1} for fd ssDNA, which we measured by diluting aliquots of stock solutions of the DNAs into the alkaline measuring solution and into aqueous buffers in which the absolute extinction coefficients for the DNAs were known (Berkowitz & Day, 1974; Wiseman & Day, 1977). Under the alkaline conditions, these single-stranded DNAs absorb only about 75% as much UV light as mixtures of their constituent free nucleotides (see the following). Since they are denatured with respect to H-bonding, the relatively strong hypochromicity apparently derives from fluctuating alignments of neighboring base chromophores.

The average value from Pf1 DNA and fd DNA is applicable to C2 DNA within its uncertainty because the extinction coefficients for the denatured states depend only weakly on base composition. For example, maximum ϵ(P) values calculated for mixtures of the free nucleotides having different base compositions of *Inovirus* species of known sequence were within 2% of each other and all occurred at 262 nm. For these calculations, we measured reference spectra for the free 5′-mononucleotides in the alkaline measuring buffer and in neutral aqueous buffers and established the extinction coefficients in the alkaline conditions with literature values for the neutral aqueous buffers (Dunn & Hall, 1970). We calculated the following values: $\epsilon(P)^{Pf1\ mix}_{262.0\ nm}$ = 10 124 M^{-1} cm^{-1}, $\epsilon(P)^{Pf3\ mix}_{262.0\ nm}$ = 9926 M^{-1} cm^{-1}, and $\epsilon(P)^{fd\ mix}_{262.0\ nm}$ = 10 111 M^{-1} cm^{-1}. Also, for the base composition of 1.1 kb of available C2 DNA sequence (18.1% A, 25.4%, C, 29.5% G, 27.0% T), we calculate $\epsilon(P)^{C2\ mix}_{262.5\ nm}$ = 10 007 M^{-1} cm^{-1}. The other base compositions are as follows: Pf1 DNA, 19.4% A, 31.3% C, 30.2% G, 19.1% T (Hill *et al.*, 1991); Pf3 DNA, 19.9% A, 21.3% C, 24.1% G, 34.8% T (Luiten *et al.*, 1985); and fd DNA, 24.6% A, 20.2% C, 20.7% G, 34.5% T (Beck *et al.*, 1978).

The stoichiometry obtained was *n/s* = 2.14 ± 0.13, which is close to the value obtained by STEM (1.96 ± 0.09). The uncertainties cited are standard deviations. There are systematic uncertainties in the two methods, such as in the reference spectra and in unknown relative moisture contents, and in addition, there are unknown contributions of minor coat proteins on the order of 2%. On the basis of the two methods, we conclude that for C2 the value of *n/s* is close to 2, but it is not necessarily an integer. Its bounds are 1.9 < *n/s* < 2.3.

UV Absorbance Spectrum. The absorption spectrum for C2 virus in 10 mM sodium phosphate (pH 7.2) from 185 to 460 nm is shown in Figure 5. The ordinate is the molar absorbance per *structure unit*, which is taken here to be one major coat protein subunit plus 2.14 ± 0.13 nucleotides; we use the spectroscopically determined *n/s* in the following, for internal consistency. The specific extinction coefficient for C2 intact virus in 10 mM sodium phosphate (pH 7.2) at λ_{max} = 264 nm is 3.51 ± 0.23 mL mg^{-1} per structure unit, without light-scattering corrections. In molar units, the

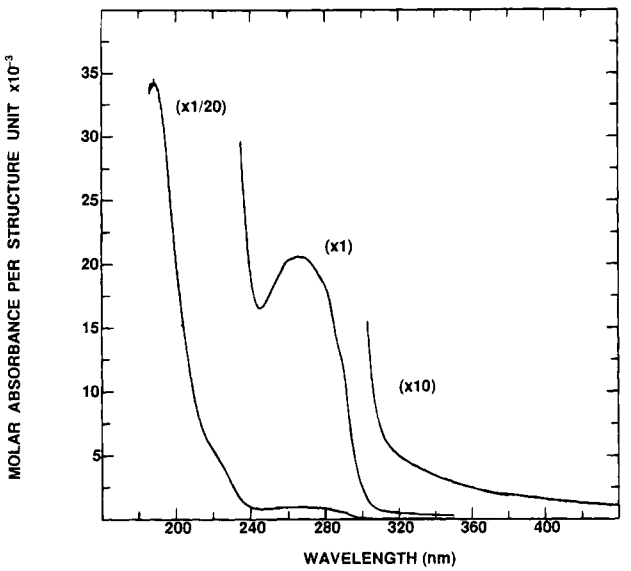

FIGURE 5: Absorbance and light scattering of C2 virus in 10 mM sodium phosphate (pH 7.2). The molar absorbance is presented per *structure unit*, taken here as 2.14 nucleotides and one subunit as obtained from spectroscopy. The molar absorbance scale is correct for the middle spectrum; ordinate values must be divided by 10 and multiplied by 20 for the longer and shorter wavelength regions, respectively.

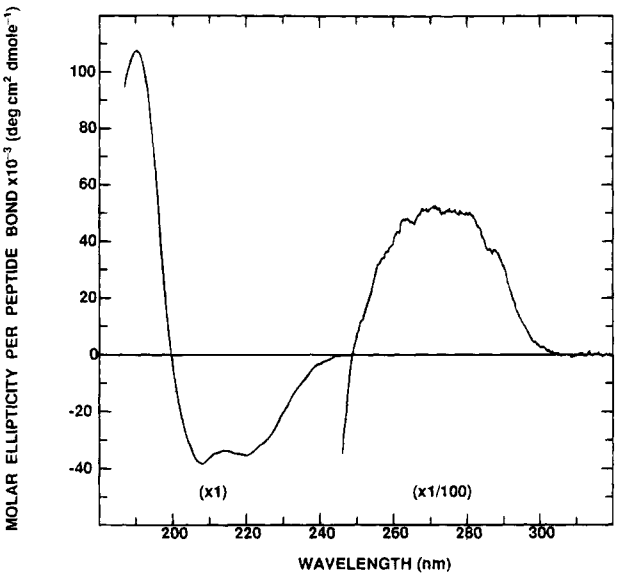

FIGURE 6: Circular dichroism of intact C2 virus in 10 mM sodium phosphate (pH 7.2). Amplitudes are expressed as molar ellipticity per average amino acid residue. Conversion to molar ellipticity per average nucleotide is done with the multiplying factor (51/2.14), which gives $[\theta]_{275\ nm}$ = 12 200 deg·cm^2 dmol^{-1} (see text).

maximum $\epsilon_{264\ nm}$ = 20 600 ± 1300 M^{-1} cm^{-1}, and the minimum $\epsilon_{245\ nm}$ = 16 500 ± 1100 M^{-1} cm^{-1}. These extinction coefficients for intact C2 virus were obtained for different samples of virus from pairs of spectra for a dilution of stock solution into a neutral buffer (10 mM sodium phosphate, pH 7.2) and a dilution into the alkaline urea solution. The coefficients were then calculated from the spectroscopically determined molar *n/s* ratio, the alkaline extinction coefficients of the components, and the molecular weights of the components. After light-scattering corrections, the extinction coefficient at 264 nm is 19 500 ± 1300 M^{-1} cm^{-1}. When the structure unit is defined as one subunit plus exactly 2 nucleotides (see above) the coefficient is 18 100 ± 1200 M^{-1} cm^{-1}. This lower value results from the way a 2/1 ratio of the separate DNA and protein spectra fit the disrupted virus spectrum of Figure 4.

Conformation of C2 DNA from Its Assigned Near-UV Absorbance and CD Spectra. Three observations with their simplest interpretations, taken together, provide a consistent indication of C2 DNA structure in the virion. The first is the assigned hypochromism. Values of ϵ(P) for double-stranded, base-stacked DNA are generally near 6500 M^{-1} cm^{-1}, and for heat-denatured, unstacked DNA they are near 9000 M^{-1} cm^{-1} [see, for example, Bloomfield *et al.* (1974) and Kostrikis *et al.* (1994)]. One can assign an *in situ* nucleotide absorbance by subtracting an estimate for the protein absorbance from the absorbance per structure unit and then dividing by *n/s*. For the protein at 260 nm, we estimate a molar absorbance of 4660 ± 250 M^{-1} cm^{-1} for the one W (3790), one Y (580), and two F's (290) from data compiled by Mihalyi (1970). The uncertainty estimate is based on the consideration that nonpolar solvent effects within proteins tend to shift spectra to longer wavelengths, which would reduce extinctions for W and Y at 260 nm, yet these same effects tend to increase the overall absorbance, which can more than compensate for any reduction at 260 nm from a shift [see, for example, Yanari and Bovey (1960) or Kostrikis *et al.* (1994)]. For C2 DNA *in situ*, the calculations yield ϵ(P) ~ 7010 ± 1550 M^{-1} cm^{-1} or ϵ(P) ~ 6590 ± 1450 M^{-1} cm^{-1}, respectively, when the higher values for the structure unit extinction and *n/s* are used or when the lower values are used. Even with the large uncertainties, these estimates of $\epsilon(P)_{260}$ are consistent with base stacking in the C2 virion.

The second observation is that the near-UV CD spectrum of C2 looks like it contains a strong contribution from a duplex DNA structure. The CD spectrum is shown in Figure 6. Protein and DNA chromophores contribute to the entire spectrum, which is presented as the molar ellipticity per amino acid residue (actually per peptide bond; see the following for an analysis of the far-UV region). To convert the amplitudes to molar ellipticities per nucleotide, one multiplies by 51 and divides by *n/s* (2.14 for the spectrum as presented in Figure 6). The spectrum above 250 nm, with its strong positive maximum at 270 nm, is similar to the spectra for classical duplex DNAs, although there are discernible fine structure shoulders at ~251, ~256, and ~262 nm, as well as near ~280 and ~288 nm, all of which indicate contributions from protein aromatics. The molar ellipticities of free L-tryptophanamide and L-tyrosinamide at 275 nm are ca. +800 and ca. −200 deg·cm^2 dmol^{-1}, respectively (Adler *et al.*, 1973); half of the sum is only ca. +300 deg·cm^2 dmol^{-1}. However, the average tyrosine in Pf1 has been shown to make a strong negative CD contribution at 275 nm: as strong as −6000 deg·cm^2 dmol^{-1} (Kostrikis *et al.*, 1994). On the other hand, tryptophan makes a strong positive contribution at 275 nm to the CD of the Ff virus and to all class I viruses of about this magnitude (L. A. Day, unpublished). This effect is accompanied by enormous effects in the far-UV CD of Ff (Arnold *et al.*, 1992), and although such far-UV CD effects are not evident in the far-UV CD of C2, we take +6000 deg·cm^2 dmol^{-1} as the maximum limit of probable contribution. Thus, adjustments to the apparent DNA amplitude from protein could be zero through cancellation, or they could be as much as ±6000 deg·cm^2 dmol^{-1}. In the face of these uncertainties, we assign

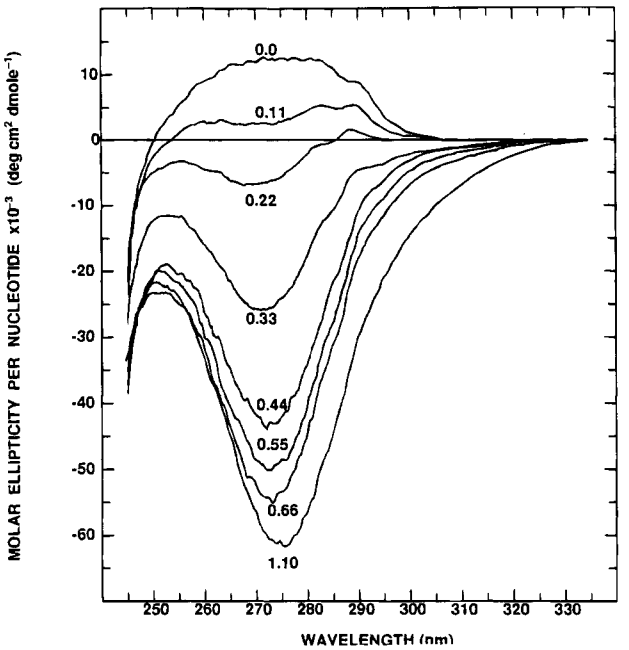

FIGURE 7: CD spectra of C2 virus and Ag−C2 complexes in 0.15 M borate (pH 8.6) from $m = 0.0$ to $m = 1.10$ (m is moles of Ag(I) per mole of nucleotide). The concentration of C2 virus was 0.279 mg mL^{-1}, corresponding to 0.032 mg mL^{-1} DNA.

$[\theta]_{275} \sim 12\,000 \pm 6000$ deg·cm^2 dmol^{-1} to the average base in C2; in alternative units, the amplitude is $\Delta\epsilon = 3.7 \pm 1.9$ M^{-1} cm^{-1} (Figures 6 and 7). This amplitude is in the range of amplitudes for right-handed duplex DNAs [for a review of the extensive DNA CD literature, see Tinoco *et al.* (1980)]. Although aromatic residues of the protein contribute in unknown ways to the CD in this region, the amplitude and the overall shape of the near-ultraviolet CD of C2 hint at a right-handed, base-stacked DNA.

The third observation is that the DNA in C2 behaves like a DNA with a base-stacked, right-handed conformation when its structure is probed with Ag(I). For some viruses, Ag(I) probing monitored by CD, according to the method introduced by Casadevall and Day (1982, 1983), can yield the chirality of the two antiparallel strands of DNA packed in virions. The chirality is established in cases for which the spectroscopic changes in response to titration mimic those in response to the titration of supercoiled dsDNA, the known chirality of which is locked in by topological linking. When Ag(I) is added to solutions of covalently closed superhelical DNA, it leads to monotonic changes in CD, ultimately producing a difference in the molar ellipticity of 55 000 deg·cm^2 dmol^{-1} at 270 nm, at a mole ratio $m = 0.50$ of Ag(I) added per nucleotide present; further additions abruptly change the type of CD spectrum generated [see Figure 2 of Casadevall and Day (1983)]. The effects are reversible. Because the right-handedness of the fundamental Watson−Crick helix of the superhelical DNA is topologically fixed, the CD changes are for right-handed DNA. Ag(I)-bridged, base-stacked, right-handed DNA produces a large negative CD band near 270 nm at $m = 0.50$.

In the case of C2 virus, titrations by Ag(I) result in changes in the CD spectrum, first at 271 nm and then shifting to longer wavelengths (Figure 7); the amplitude change at $m = 0.50$ is about 60 000 deg·cm^2 dmol^{-1}. The total change at 275 nm is $-74\,000 \pm 4000$ deg·cm^2 dmol^{-1}. The CD spectral changes produced by adding Ag(I) were reversed by dialysis against 0.1 M NaCl. CD changes below 250 nm were negligible. At the beginning of the C2 titration, the changes mimic, almost exactly, those observed by Casadevall and Day (1982, 1983) for supercoiled DNA. The most direct interpretation of the present results is that the two antiparallel DNA strands in C2 wind around each other in a right-handed fashion, with the bases stacked and directed toward each other so that Ag(I) bridges can form between bases in opposite strands.

Thus, a base-stacked, right-handed DNA conformation, with a rise of 3.2 Å, is indicated by the combination of three spectroscopic observations, the virion length, and the DNA size. For such a DNA, the molar ellipticity estimate of $12\,000 \pm 6000$ deg·cm^2 dmol^{-1}, at face value, corresponds to a rotation per residue in the range 30−40°, according to the calculations of Johnson *et al.* (1980). This range of rotation anges and the rise per residue of 3.2 Å suggest a pitch in the range of 29−38 Å.

Far-Ultraviolet CD. Below 240 nm, the CD spectrum gives information about protein conformation. The number of nucleotide chromophores in C2 virus is ~25 times lower than the number of peptide chromophores, and their molar ellipticities are lower than those of helical peptides. For the C2 spectrum below 240 nm in Figure 6, free native DNA and free denatured DNA have CD amplitudes on the order of 10^{-2} those for the virus. In terms of protein, the amplitudes at key wavelengths below 240 nm are $[\theta]_{190\ nm} = +108\,000$, $[\theta]_{208\ nm} = -38\,500$, and $[\theta]_{220\ nm} = -35\,500$ deg·cm^2 dmol^{-1}, from calculations done with $n/s = 2.14$. The CD amplitudes are 7% lower when calculations are carried through with $n/s = 2$. The shape of the spectrum is like those of α-helix reference spectra, but the crossover is at 199.5 nm, a lower wavelength than for many reference spectra, and the 208 nm band is larger in amplitude than the 220 nm band. The amplitudes are among the highest observed for helical proteins and polypeptides (Greenfield & Fasman, 1969; Sreerama & Woody, 1993). The combined features indicate a largely α-helical conformation, but, as discussed in the following, some 3_{10}-helix may be present.

DISCUSSION

Our interest in C2 virus structure derives from its relation to the structures of the other filamentous viruses, and here we consider the presently available C2 results in the context of other viruses, primarily fd, If1, and IKe (class I) and Pf1, Xf, and Pf3 (class II). We begin with comparisons of macroscopic properties from the electron micrographs of Bradley *et al.* (1982), which first showed that one end of the virion is tapered and the other end has a recognizable disklike structure not seen for the other viruses. In two different micrographs, there were exactly 19 virions lying parallel to one another in several regions along their contours, yet aggregated at other points and terminating at the same place, as if bundles of 19 virions were unraveling. Aggregation effects seem to be strong in concentrated C2 solutions and may produce groups of 7 (1 surround by 6) and groups of 19 (a group of 7 surrounded by 12 more). Images of individual virion STEM grids (Figure 3) appear to be wavier than images of most of the other viruses, although images of *Inovirus* species X are wavier than those of C2 (Bradley *et al.*, 1981). C2 has an apparent persistance length that is much lower than those of the known class I

and class II viruses (S. A. Reisberg, unpublished work). Furthermore, C2 virus particles bind to the sides of the pili of the host cells, as does Pf3 (Bradley & Whelan, 1989), rather than at the ends, as is the case for Ff and Pf1 (Caro & Schnoss, 1966; Bradley, 1973), or on the tapered sides of the tips, as is the case for tf-1 (Coetzee *et al.*, 1987).

Helical Protein Conformation. The C2 coat protein is highly helical, in spite of the fact that the first five residues cannot be in an α-helix or in a 3_{10}-helix, because of two prolines (P_3 and P_6). Even so, the amplitudes are extremely high: as high or higher than amplitudes of polypeptides in α-helical conformations (Holzwarth *et al.*, 1962; Holzwarth & Doty, 1965; Carver *et al.*, 1966; Greenfield & Fasman, 1969) and much higher than amplitudes for shorter α-helical segments in globular proteins deduced from analyses of CD spectra and X-ray structural data (Chen *et al.*, 1974; Chen & Yang, 1977; Brahms & Brahms, 1980; Johnson, 1988; Sreerama & Woody, 1993). With the five N-terminal residues being nonhelical, one might view the CD spectrum as characterizing the conformation of the remaining 47 residues, in which case it is quite similar to the far-ultraviolet CD spectrum of Pf1 virus (Kostrikis *et al.*, 1994). Like the Pf1 spectrum, the C2 spectrum is blue shifted relative to the CD spectra of class I viruses and the Pf3 virus (Day *et al.*, 1988; Clack & Gray, 1989; Arnold *et al.*, 1992), having its crossover from negative to positive ellipticity at 199.5 nm, rather than 201−203 nm, and its 208 nm minimum is deeper than its 220 nm minimum. On the basis of CD spectra recently obtained theoretically and experimentally for 3_{10}-helices (Manning & Woody, 1991; Miick *et al.*, 1992), these features hint at the possible presence of some 3_{10}-helix in C2. The considerations are explained in more detail by Kostrikis *et al.* (1994) for the case of Pf1. We conclude that the C2 coat protein in the virion is mostly helical, with perhaps some 3_{10}-helix.

From the similarities in overall C2 life cycle and in the sequence pattern of its coat protein compared to other *Inovirus* species, it seems highly likely that C2 coat protein is an intrinsic membrane protein prior to being assembled into the virion. Kuhn and co-workers have shown that Pf3 protein acts as its own "reverse signal peptide", in that the positive C-terminal region and its contiguous hydrophobic domain dictate the membrane insertion, instead of a positive N-terminal region and the contiguous hydrophobic domain from the usual signal peptide (Kuhn *et al.*, 1990; Rohrer & Kuhn, 1990; Lee *et al.*, 1992; Thiaudiere *et al.*, 1993). C2 and Pf3 coat proteins are so similar to each other in having only two negatively charged residues and no signal peptide that it is plausible that C2 protein also functions as its own reverse signal peptide for membrane insertion. Although we cannot predict the structure the C2 protein would have in the membrane of the infected cell, it is likely that it would differ significantly from its highly helical structure in the virion.

DNA Interaction Domain of the Protein. In the C-terminal DNA interaction domain there are three basic residues, whereas in the other viruses there two or four basic residues (Figure 1). Their spacing is i, $i + 7$, $i + 10$, so that, given an α-helical conformation, the three side chains can point in approximately the same direction, presumably inward toward the DNA. The class I viruses all have four basic residues in strictly conserved positions in their DNA interaction domains; class II virus Xf also has four in this domain. There are mutants of Ff(fd) having three basic residues per subunit, not four, in their DNA interaction domains, mutants in which K_{48} was replaced by neutral residues (Hunter *et al.*, 1987). The mutant virions are 1.33−1.38 times longer than wild type, from which one can deduce *n/s* values near 1.8 (instead of the 2.4 for fd wild type) and an average axial rise per nucleotide near 3.6 Å (instead of the 2.7 Å for fd wild type). Even though they each have three basic residues to interact with DNA, C2 and these mutants of fd clearly differ with respect to *n/s* and average rise per nucleotide.

Helical Symmetries, Amino Acid Sequences, and DNA/Protein Stoichiometries. Inovirus structure classes are defined by the helical symmetries of the protein coats, as revealed by X-ray fiber diffraction. So far, only two classes have been observed, although six viruses have been investigated (Marvin *et al.*, 1974a,b; Makowski & Caspar, 1981; Peterson *et al.*, 1982). The class I and class II symmetries are similar in that both are intrinsic to the close-packing of α-helical subunits around two strands of DNA, but they are so fundamentally different that it seems unlikely that viruses can move from one symmetry class to the other through mutations. Theoretical work has shown that these two symmetries are among a small number of optimal symmetries in relation to all sterically possible close-packed arrangements of α-helical coat protein subunits in such structures (Marzec & Day, 1988). Class II symmetry, in addition to being stereochemically optimal, appears to allow optimal electrostatic interactions within the virions of at least two species, Xf and Pf1 (Marzec & Day, 1994). A key aspect of filamentous virus structure is the relation between the protein and DNA structures, but evidence on DNA helical symmetry from X-ray fiber diffraction is only available for fd (Banner *et al.*, 1981; Marzec & Day, 1983) and Pf1 (Bryan *et al.*, 1983; Liu & Day, 1994). In the absence of a direct determination of C2 protein and DNA symmetry by diffraction measurements, it is of interest to consider the C2 protein sequence and the C2 DNA/protein stoichiometry in comparison to those of the six viruses in the two observed symmetry classes.

The sequences are shown in Figure 1 with their C-terminal DNA interaction domains aligned, and detailed sequence comparisons are shown in Table 3 and Figure 8 for some pairs of sequences. The class I viruses fd, IKe, and If1 have closely homologous sequences and similar sizes of 50−53 amino acids. The sequences all have strictly conserved positions in their DNA interaction domains and contiguous hydrophobic domains. Counting from the C-termini, one sees basic amino acids at positions −2, −6, −7, and −10, two F residues at positions −5 and −8, and a key W residue at position −25. It would appear that the shared features are part of an natural optimization, over the course of evolution, of the class I packing arrangement. I2-2, which contains 55 amino acids, shares these conserved sequence features with all class I viruses and has limited homology to class II viruses. I2-2 therefore is placed in class I solely on the basis of its sequence. The inference of symmetry from the sequence for I2-2 seems well based because of the high homology. By contrast, C2 shows little overall homology to any class I protein (Table 3, Figure 8), and it has none of the strictly conserved features of class I sequences (Figure 1). Thus, there is no suggestion from sequence that C2 subunits would pack together with class I helical symmetry.

Table 3: Statistics from Representative Pairwise Comparisons of Major Coat Protein Sequences[a]

groups	pairs	quality of fit	length	quality ratio	gaps	percent similarity	percent identity
	C2−C2	78.0	52	1.500	0	100.0	100.0
C2 *vs* I	C2−fd	21.9	46	0.521	1	42.9	16.7
	C2−If1	20.4	51	0.443	4	59.5	33.3
	C2−IKe	23.7	52	0.551	2	55.8	27.9
	C2−I22	21.0	52	0.488	1	41.9	23.3
C2 *vs* II	C2−Pf1	23.8	48	0.517	1	47.8	19.6
	C2−Pf3	22.2	45	0.516	2	53.7	29.3
	C2−Xf	28.8	48	0.626	1	58.7	26.1
fd *vs* I	fd−I22	39.4	54	0.804	1	67.4	40.9
	fd-If1	58.6	50	1.172	0	82.0	68.0
	fd−IKe	40.4	52	0.824	1	65.3	44.9
fd *vs* II	fd−Pf1	22.4	42	0.533	0	45.2	16.7
	fd−Pf3	20.5	43	0.526	1	48.7	20.5
	fd−Xf	23.0	49	0.523	2	56.8	25.0
Pf1 *vs* I	Pf1−I22	26.6	49	0.591	1	55.6	24.4
	Pf1−If1	24.5	45	0.544	0	48.9	17.8
	Pf1−IKe	23.7	46	0.564	1	54.8	19.0
Pf1 *vs* II	Pf1−Pf3	26.2	42	0.672	1	59.0	33.3
	Pf1−Xf	30.5	46	0.726	1	64.3	38.1
	Pf3−Xf	25.4	39	0.651	0	53.8	20.5

[a] Results of calculations using the local homology algorithms described by Needleman and Wunsch (1970) and Smith and Waterman (1981) as implemented in the program BESTFIT by P. Haeberli in the GCG package (Version 7.3, 1991) of sequence analysis programs.

The C2 sequence shows slightly better fits to class II sequences as a group than to class I virus sequences as a group. C2 and Xf show 12 identities at 28 positions over their C-terminal portions when a single gap of 4 residues is inserted. Pf1 and Pf3 also show some homology to the C2 sequence when a single gap of four residues is inserted (Figure 8). The class II viruses Pf1, Xf, and Pf3 are heterodisperse with respect to sequence, but they are all close in size, 44−46 amino acids, and it may be significant that the gap size that gives suggestions of homology for all three is four residues. Thus, were C2 to be classified solely on the basis of its sequence into either of the two known classes, the choice would be class II. This is primarily because of the connection to Xf.

The defining characteristic of a class II protein helix is its rotation angle of ~60° per subunit, corresponding to ~5.4 subunits per turn. Were C2 to have ~5.4 subunits per turn, the rise per subunit of ~3.2 Å would give a pitch for the C2 subunit helix of about 17 Å. We have deduced, however, that the DNA helix in C2 may well have a pitch in the range of 33 ± 5 Å. Thus, if C2 is in class II, it appears that it would be a class II virus, with different pitches for its protein subunit helix and its DNA helix. There is a precedent for this type of *Inovirus* in that Xf has a pitch of about 16 Å for its subunit helix and a pitch of about 30 Å for its DNA helix (Casadevall & Day, 1983; Day *et al.*, 1988). Both C2 and Xf have *n/s* values near 2, and their near-UV CD spectra are very similar.

```
C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
If1       ADDATSQAKAAFDSLTAQATEMSGYAWALVVLVVGATVGIKLFKKFVSRAS

C2      MGPTAPTDIASL.ASSVDFSSV..GLGILAIAGTV.ITLYVTWKGAQFVIRQVRGA
If1   ADDATSQAKAAFDSLTAQATEMSGYAWALVVLVVGATVGIKLF.....KKFVSRAS

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYV.TWKGAQFVIRQVRGA
If1     ADDATSQAKAAFDSLTAQATEMSGYAWALVVLVVGATVGIKLFKKFVSRAS

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf1           GVIDTSAVQSAITDGQGDMKAIGGYIVGALVILAVAGLIYSMLRKA

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf1         GVIDTSAVQSAITDGQGDMK..AIGGYIVGALVILAVAGLIYSMLRKA

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf1       GVIDTSAVQSAITDGQGDMKAIGGYIVGALVIL....AVAGLIYSMLRKA

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf3               MQSVITDVTGQLTAVQADITTIGGAIIVLAAVVLGIRWIKAQFF

C2      MGPTAPTDIASLASSV..DFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf3       MQSVITDVTGQLTAVQADITTIGGAIIVLAAVVLG..IRWIKAQFF

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Pf3           MQSVITDVTGQLTAVQADITTIGGAIIVL....AAVVLGIRWIKAQFF

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Xf              SGGGGVDVGDVVSAIQGAAGPIAAIGGAVLTVMVGIKVYKWVRRAM

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Xf        SGGGGVDVGDVVSAIQ..GAAGPIAAIGGAVLTVMVGIKVYKWVRRAM

C2      MGPTAPTDIASLASSVDFSSVGLGILAIAGTVITLYVTWKGAQFVIRQVRGA
Xf           SGGGGVDVGDVVSAIQGAAGPIAAIGGAVLTVMVGIK....VYKWVRRAM
```

FIGURE 8: C2 coat protein sequence compared to the sequence of If1, as representative of class I viruses and compared to those of all three class II viruses. The results of three modes of pairwise comparison are shown. First is the best match with no gaps introduced and the C-termini aligned within ±3 residues, second is the computed best fit (see Table 3), and third is the best match within the basic DNA interaction domain and its contiguous hydrophobic domain, with one gap introduced. These domains are directly involved in DNA packing and in a bend region needed for the overlapping of subunits in the virion; the N-terminal regions are on the outside of the structure and show the most sequence divergence (see also Figure 1).

Nevertheless, it is possible that C2 is in neither class I nor class II, even though these are the only two symmetries yet observed. There are a few sterically feasible ways to pack α-helical subunits in addition to the class I and class II symmetries that are predicted by the close-packing theory of Marzec and Day (1988). Some of these additional symmetry classes call for pitches in the 33 ± 5 Å range for the C2 protein helix, so that the DNA and protein parts in C2 could have the same pitch. In fact, for *n/s* exactly 2, it would even be possible for the loop of DNA to have exactly the same symmetry as the protein helix by having a structure unit consisting of one nucleotide from one strand, one nucleotide from the other strand, and one subunit.

Whether C2 structure defines a new symmetry class or is an unusual member of class II will be decided when diffraction data are available for interpretation in conjunction with the data presented here. In any case, the further study of C2 is expected to advance our understanding of *Inovirus* architecture.

ACKNOWLEDGMENT

The other authors express their sadness over the deaths of coauthor Steve Reisberg and his wife Jenny Sun, who died on a hiking trip in September, 1990. Dr. Reisberg's micrographs and analyses provided early indication that C2 is different from the more well characterized strains. The help of Joseph S. Wall, Martha N. Simon, and Beth Lin of the Brookhaven STEM facility was essential to this study and is gratefully acknowledged. Help from D. J. Liu in digitizing some of the spectra, Yang-gu Shi and Y.-K. Yip in protein chemistry, and Chris Marzec and Richard Hendrix through discussions is also gratefully acknowledged.

REFERENCES

Adler, J. T., Greenfield, N. J., & Fasman, G. D. (1973) *Methods Enzymol. 27*, 675−735.
Anderer, F. A. (1959) *Z. Naturforsch. 14b*, 24−28.
Arnold, G. E., Day, L. A., & Dunker, A. K. (1992) *Biochemistry 31*, 7948−7956.
Banner, D., Nave, C., & Marvin, D. A. (1981) *Nature 289*, 814−816.
Beck, E., Sommer, R., Auerswald, E. A., Kurz, C., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T., & Takanami, M. (1978) *Nucleic Acids Res. 5*, 4495−4503.
Berkowitz, S. A., & Day, L. A. (1974) *Biochemistry 13*, 4825−4831.
Bloomfield, V. A., Crothers, D. M., & Tinoco, I. (1974) *Physical Chemistry of Nucleic Acids*, Harper & Tow, New York.
Bradley, D. E. (1973) *Can. J. Microbiol. 19*, 623−632.
Bradley, D. E., & Whelan, J. (1989) *J. Gen. Microbiol. 135*, 1857−1863.
Bradley, D. E., Coetzee, J. N., Bothma, T., & Hedges, R. W. (1981) *J. Gen. Microbiol. 126*, 389−396.
Bradley, D. E., Sirgel, F. A., Coetzee, J. N., Hedges, R. W., & Coetzee, W. F. (1982) *J. Gen. Microbiol. 128*, 2485−2498.
Brahms, S., & Brahms, J. (1980) *J. Mol. Biol. 138*, 149−178.
Bryan, R. K., Bansal, M., Folkhard, W., Nave, C., & Marvin, D. A. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4728−4731.
Carne, A., Hill, D. F., Stockwell, P. A., Hughes, G., & Petersen, G. B. (1991) *Proc. R. Soc. London 245*, 23−30.
Caro, L. G., & Schnos, M. (1966) *Proc. Natl. Acad. Sci. U.S.A. 56*, 126−132.
Carver, J. P., Schechter, E., & Blout, E. R. (1966) *J. Am. Chem. Soc. 88*, 2550−2561.
Casadevall, A., & Day, L. A. (1982) *Nucleic Acids Res. 10*, 2467−2481.
Casadevall, A., & Day, L. A. (1983) *Biochemistry 22*, 4831−4842.
Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10* (14), 1195−1207.
Chen, Y., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350−3359.
Clack, B. A., & Gray, D. M. (1989) *Biopolymers 28*, 1861−1873.
Coetzee, J. N., Bradley, D. E., Hedges, R. W., Tweehuizen, M., & Du Toit, L. (1987) *J. Gen. Microbiol. 133*, 953−960.
Day, L. A. (1969) *J. Mol. Biol. 39*, 265−277.
Day, L. A., & Hoppensteadt, F. (1972) *Biopolymers 11*, 2131−2140.
Day, L. A., Marzec, C. J., Reisberg, S. A., & Casadevall, A. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 509−539.
Dunn, D. B., & Hall, R. H. (1970) in *CRC Handbook of Biochemistry*, 2nd ed. (Sober, H. A., Ed.) pp G3−98, CRC, Cleveland, OH
Frangione, B., Nakashima, Y., Konigsberg, W., & Wiseman, R. L. (1978) *FEBS Lett. 96*, 381−384.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
Greenfield, N. J., & Fasman, G. D. (1969) *Biochemistry 8*, 4108−4116.
Greenwood, J., & Perham, R. N. (1989) *Virology 171*, 444−452.
Guy-Caffey, J. K., Rapoza, M. P., Jolley, K. A., & Webster, R. E. (1992) *J. Bacteriol. 174*, 260−265.
Hill, D. F., & Petersen, G. B. (1982) *J. Virol. 44*, 32−46.
Hill, D. F., Short, N. J., Perham, R. N., & Petersen, G. B. (1991) *J. Mol. Biol. 218*, 349−364.
Holzwarth, G., & Doty, P. (1965) *J. Am. Chem. Soc. 87*, 218−228.
Holzwarth, G., Gratzer, W. B., & Doty, P. (1962) *J. Am. Chem. Soc. 84*, 3194−3196.
Hunter, G. J., Rowitch, D. H., & Perham, R. N. (1987) *Nature 327*, 252−254.
Johnson, W. C., Jr. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 145−166.
Johnson, B. B., Dahl, K. S., Tinoco, I., Ivanov, V. L., & Zhurkin, V. B. (1981) *Biochemistry 20*, 73−78.
Kazmierczak, B. I., Mielke, D. L., Russel, M., & Model, P. (1994) *J. Mol. Biol. 238*, 187−98.
Knippers, R., & Hoffmann-Berling, H. (1966) *J. Mol. Biol. 21*, 305−312.
Kostrikis, L. G. (1993) Architectural studies on filamentous bacterial viruses: Pf1, Epf1, and C2, Ph.D. Thesis, New York University, New York.
Kostrikis, L. G., Liu, D. J., & Day, L. A. (1994) *Biochemistry 33*, 1694−1703.
Kuhn, A., Kreil, G., & Wickner, W. (1987) *EMBO J. 6*, 501−505.
Kuhn, A., Rohrer, J., & Gallusser, A. (1990) *J. Struct. Biol. 104*, 38−43.
Lee, J., Kuhn, A., & Dalbey, R. E. (1992) *J. Biol. Chem. 267*, 938−943.
Liu, D. J., & Day, L. A. (1994) *Science 265*, 671−675.
Luiten, R. G. M., Schoenmakers, J. G. G., & Konings, R. N. H. (1983) *Nucleic Acids Res. 11*, 8073−8085.
Luiten, R. G. M., Putterman, D. G., Schoenmakers, J. G. G., Konings, R. N. H., & Day, L. A. (1985) *J. Virol. 56*, 268−276.
Makowski, L., & Caspar, D. L. D. (1981) *J. Mol. Biol. 145*, 611−617.
Makowski, L., Caspar, D. L. D., & Marvin, D. A. (1980) *J. Mol. Biol. 140*, 149−181.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning—A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Manning, M. C., & Woody, R. W. (1991) *Biopolymers 31*, 569−586.
Marvin, D. A., Wiseman, R. L., & Wachtel, E. (1974) *J. Mol. Biol. 82*, 121−138.
Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J., & Marvin, F. J. (1974) *J. Mol. Biol. 88*, 581−600.
Marzec, C. J., & Day, L. A. (1983) *Biophys. J. 42*, 171−180.
Marzec, C. J., & Day, L. A. (1988) *Biophys. J. 53*, 425−440.
Marzec, C. J., & Day, L. A. (1994) *Biophys. J. 67*, 2205−2222.
Mihalyi, E. (1970) in *CRC Handbook of Biochemistry*, 2nd ed. (Sober, H. A., Ed.) pp B75−77, CRC, Cleveland, OH.
Miick, S. M., Martinez, G. V., Fiori, W. R., Todd, A. P., & Millhauser, G. L. (1992) *Nature 359*, 653−655.

Nave, C., Fowler, A. G., Ladner, J. E., Marvin, D. A., Provencher, S. W., Tsugita, A., Armstrong, J., & Perham, R. N. (1981) *J. Mol. Biol. 149*, 675−707.
Needleman, S. B., & Wunsch, C. D. (1970) *J. Mol. Biol. 48*, 443−53.
Newman, J., Swinney, H. L., & Day, L. A. (1977) *J. Mol. Biol. 116*, 593−603.
Newman, J., Day, L. A., Dalack, G., & Eden, D. (1982) *Biochemistry 21*, 3352−3358.
Peeters, B. P., Peters, R. M., Schoenmakers, J. G., & Konings, R. N. (1985) *J. Mol. Biol. 181*, 27−39.
Peterson, C., Winter, W. T., Dalack, G., & Day, L. A. (1982) *J. Mol. Biol. 162*, 877−881.
Putterman, D. G. (1983) The filamentous bacteriophage Pf3: an investigation of its genomic sequence, Ph.D. Thesis, New York University, New York.
Putterman, D. G., Casadevall, A., Boyle, P. D., Yang, H. L., Frangione, B., & Day, L. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 88*, 699−703.
Radloff, R., Bauer, W., & Vinograd, J. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1514−1521.
Reisberg, S. A. (1989) Radial mass density profiles, mass per length, and nucleotide/subunit ratios for filamentous bacteriophages from electron microscopy, Ph.D. Thesis, New York University, New York.
Rohrer, J., & Kuhn, A. (1990) *Science 250*, 1418−1421.
Russel, M. (1991) *Mol. Microbiol. 5*, 1607−1613.
Russel, M. (1993) *J. Mol. Biol. 231*, 689−697.
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161−178.
Shen, L. M., Lee, J. I., Cheng, S. Y., Jutte, H., Kuhn, A., & Dalbey, R. E. (1991) *Biochemistry 30*, 11775−81.
Shin, S., & Day, L. A. (1995) *Anal. Biochem.* (in press).
Simons, G. F. M., Konings, R. N. H., & Schoenmakers, J. G. G. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4194−4198.
Smilowitz, H. (1974) *J. Virol. 13*, 94−99.
Smith, T. F., & Waterman, M. S. (1981) *Adv. Appl. Math. 2*, 482−89.
Snell, D. T., & Offord, R. E. (1972) *Biochem. J. 127*, 167−173.
Sreerama, N., & Woody, R. W. (1993) *Anal. Biochem. 209*, 32−44.
Stassen, A. P. M., Schoenmakers, E. F. P. M., Yu, M., Schoenmakers, J. G. G., & Konings, R. N. H. (1992) *J. Mol. Evol. 34*, 141−152.
Thiaudiere, E., Soekarjo, M., Kuchinka, E., Kuhn, A., & Vogel, H. (1993) *Biochemistry 32*, 12186−96.
Tinoco, I., Jr., Bustamante, C., & Maestre, M. F. (1980) *Annu. Rev. Biophys. Bioeng. 9*, 107−141.
Van Wezenbeek, P. M. G. F., Hulsebos, T. J. M., & Schoenmakers, J. G. G. (1980) *Gene 11*, 129−148.
Wall, J. S., & Hainfeld, J. F. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 355−376.
Wiseman, R. L., & Day, L. A. (1977) *J. Mol. Biol. 116*, 607−611.
Yamamoto, K. R., Alberts, B. M., Benzinger, R., & Lawhorne, L. (1970) *Virology 40*, 734−744.
Yanari, S., & Bovey, F. A. (1960) *J. Biol. Chem. 235*, 2818−2826.

BI942060L