Circular Dichroism and Ultraviolet Absorption of a Deoxyribonucleic Acid Binding Protein of Filamentous Bacteriophage†

Loren A. Day

ABSTRACT: Gene 5 (G5) protein of fd bacteriophage has an unusual circular dichroic (CD) spectrum with a positive band at 228 nm. This band has been assigned to tyrosine according to its pH dependence and data for model compounds. Computer fitting to reference CD spectra for α-helical, β, and random conformations, as well as for tyrosyl contributions, show that the CD of G5 protein is that of a non-α-helical protein. Binding to single-stranded DNA leads to little change in average peptide conformation, although changes in the tyrosyl CD band and in tyrosyl difference absorbance bands are observed. Changes in absorbance intensity over a broad region around 262 nm show that the bases of single-stranded DNA in complexes with G5 protein are unstacked regardless of their prior state in isolated DNA. The various spectral changes have been used to show that each protein monomer of 10,000 daltons binds four nucleotides. The fd DNA–G5 protein complex, having unstacked bases and non-α-helical subunits, is very different from the fd bacteriophage particle, which has stacked bases and highly α-helical subunits of G8 protein. It is suggested that circular single-stranded fd DNA is shifted from an unstacked to a stacked state by means of a subunit exchange reaction during the course of replication and phage assembly.

The purpose of this study was to characterize the interaction of single-stranded DNA from bacteriophage fd with one of the proteins which is coded by it. This protein, the product of gene 5 of the phage genome, hereafter called G5 protein, is involved in the replication of fd DNA (Pratt and Erdahl, 1968) presumably by functioning as a repressor of complementary strand synthesis at later stages of infection (Salstrom and Pratt, 1971) and by performing some positive function in the synthesis of single-stranded DNA (Staudenbauer and Hofschneider, 1973). Oey and Knippers (1972) and Alberts *et al.* (1972) have shown that 10^5 molecules of G5 protein are present in the cell during wild-type infection, that its molecular weight is close to 10,000 daltons by gel electrophoresis and sedimentation velocity, and that it binds to single-stranded DNA in a cooperative manner. G5 protein is not present in the filamentous virion (Henry and Pratt, 1969) in which the single-strand DNA of 2×10^6 daltons is bound by about 3000 subunits of a major coat protein, each weighing 5169 daltons (Asbeck *et al.*, 1969) and coded by gene 8 (Pratt *et al.*, 1969) and three or four molecules of a minor protein coded by gene 3 (Henry and Pratt, 1969; Beaudoin, 1970).

It seems reasonable, from evidence in the references cited above, that newly replicated single-stranded viral DNA molecules are present as complexes with approximately 1500 subunits of G5 protein in fd-infected cells, and that the G5 protein subunits are replaced with subunits of the major coat protein in the assembly of progeny virions. Spectroscopic evidence is presented below on conformational aspects of such a reaction, as well as on the stoichiometry of the *in vitro* G5 protein–DNA interaction and on the involvement of tyrosine in this interaction.

Materials and Methods

Several preparations of G5 protein were used. Preparations G5A, G5B, and G5D were received from Bruce M. Alberts and A. Keith Dunker as frozen solutions, and preparation G5C was received lyophilized from Rolf Knippers. All other preparations were isolated from *Escherichia coli* K12 3300 infected with wild-type fd bacteriophage by a method based on that of Alberts *et al.* (1972), with some modifications suggested by A. Keith Dunker. Infected cells were stored frozen at $-20°$, thawed, washed with buffer I (10 mM $MgCl_2$, 2 mM $CaCl_2$, 1 mM β-mercaptoethanol, and 1 mM EDTA in 0.02 M Tris-HCl, pH 8.1), and disrupted with glass beads in a high-speed shaker. The extract, approximately 20 ml/g of wet cells, was incubated for 3 hr at 4° in the presence of 20 μg/ml of pancreatic ribonuclease (Worthington Biochemical, lot 693) and 20 μg/ml of pancreatic deoxyribonuclease 1 (Worthington Biochemical, lot D1227-30). It was then made 1 M in NaCl, incubated for 1 more hr, and centrifuged at 10,000*g* for 20 min. To each 5.5 ml of cold extract solution was added 4.5 ml of saturated $(NH_4)_2SO_4$ at room temperature. Precipitate was allowed to develop for 1 hr in an ice bath, spun down at 10,000*g* for 20 min, and resuspended in approximately 1 ml of cold buffer II (1 mM β-mercaptoethanol, 5 mM EDTA, 0.05 M NaCl, and 10% v/v glycerol in 0.02 M Tris-HCl, pH 8.1) per g of wet cells and dialyzed overnight at 4° against buffer II. After dialysis, the sample was clarified at 30,000*g* for 3 hr and applied to a DNA-cellulose column (Alberts and Herrick, 1970) equilibrated with buffer II. Salmon sperm DNA (A grade, Calbiochem, lot 010021) and Whatman CF11 cellulose were used to prepare the DNA-cellulose. A single column measuring 2.5 × 10 cm maintained at 4° was used for several G5 protein preparations. After sample application and washing with buffer II, the column was eluted by steps with buffer II containing 0.4, 0.8, and 2 M NaCl. Except for a few leading and trailing fractions, the material eluted with 0.8 M NaCl–buffer II was homogeneous by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate and

† From the Department of Biochemistry, The Public Health Research Institute of the City of New York, Inc., New York, New York 10016. *Received October 31, 1972.* Supported by Grant No. AI-09049 and 1 K04 GM-70363 from the National Institutes of Health, U. S. Public Health Service.

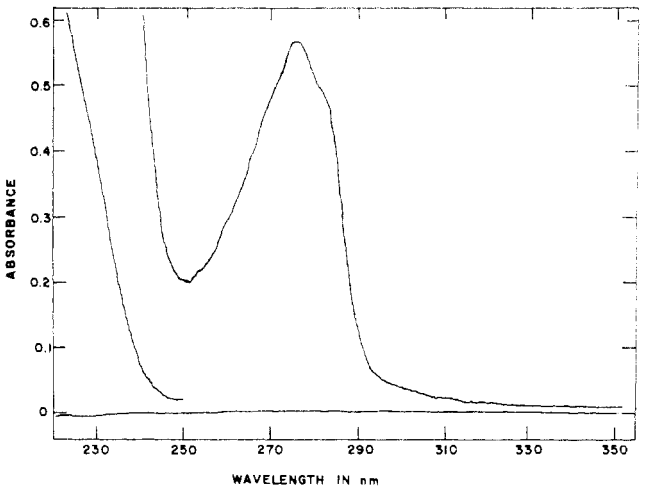

FIGURE 1: Absorbance spectra for a solution of 0.74 mg/ml of G5-protein in 0.15 M $NaClO_4$–0.02 M sodium phosphate (pH 7) with light paths of 10 mm and (350–240 nm) and 1 mm (250–225 nm).

was taken as purified gene 5 protein. The yields were approximately 1 mg of G5 protein/g of original wet cells. The yields were about the same with log- and stationary-phase wild-type infected cells. It was observed that protein samples taken directly from the DNA-cellulose columns and diluted or passed rapidly over Bio-Gel P4 columns for buffer exchange scattered less light and bound more DNA than samples prepared by dialysis and lyophilization. Storage at 4° for several days and/or freezing of DNA-cellulose column fractions in 0.8 M NaCl–buffer II led to little change in DNA binding capacity.

Protein concentrations were routinely determined with an absorbance coefficient obtained in the present study. Concentration measurements for the absorbance coefficient were made with a Brice-Phoenix differential refractometer. Protein solutions were dialyzed at 4° for 2 days prior to measurement, the final several hours of which were at room temperature to allow complete equilibration of diffusible solutes. The outer dialysate was used in the reference compartment of the differential refractometer and in the reference cell of the spectrophotometer in subsequent absorbance measurements. A refractive increment, dn/dc, at 546 nm of 0.185 ml g^{-1} was assumed; of over 30 values of dn/dc reported for more than 15 simple proteins (*cf.* Timasheff, 1970) under solvent conditions comparable to those used for G5 protein, approximately 80% are within ±3% of 0.185 ml g^{-1} and all values are within ±5%.

The fd DNA was isolated from purified fd bacteriophage by a phenol method based on that of Gierer and Schramm (1956) for RNA. Double-stranded form I DNA of PM2 bacteriophage (Espejo and Canello, 1968) was a gift from R. D. Camerini-Otero. It was better than 95% pure form I as assayed by ethidium bromide–CsCl equilibrium density gradient centrifugation and by sucrose density gradient velocity sedimentation.

Concentrations of single-stranded fd DNA solutions were obtained from extinction coefficients determined previously for fd DNA in defined buffers (Hoffmann-Berling *et al.*, 1963; Day, 1969; Schaller *et al.*, 1969). Based on the agreement between the various reported values, the uncertainty in DNA concentrations is considered to be ±3%.

An Agla micrometer syringe drive (Burroughs-Welcome, Ltd. Inc.) in combination with 250-μl Hamilton syringe fitted with a short Teflon capillary delivery tube permitted reproducible delivery of titrants to within ±0.2 μl. The absolute titrant volumes were established to within ±1% from the weight of distilled water delivered by the micrometer–syringe unit.

TABLE I: Extinction Coefficient of G5 Protein in 0.15 M $NaClO_4$–0.02 M Sodium Phosphate (pH 7.0).

Sample	$10^3 \times \Delta n$[a]	Concn[b] (mg/ml)	OD at 276[c] (1-cm Path)	$A_{1\,cm}^{1\%}$
G5A	0.081 ± 0.003	0.438	0.317	7.24
G5B	0.143 ± 0.003	0.773	0.567	7.34
G5C	0.114 ± 0.003	0.613	0.446	7.28

[a] Six readings were made with protein solution in one compartment of the refractometer cell and outer dialysate in the other. Six readings were made with outer dialysate in both compartments before and after the solution readings were taken. The errors indicated are the standard deviations of the readings; the standard deviations of the means were all ±0.001. [b] Concentrations were obtained from Δn with the assumption that dn/dc = 0.185 ml g^{-1}. [c] Absorbance values given have been corrected for light scattering. The scattering corrections were 18, 6, and 5% of the uncorrected amplitudes at 276 nm for the samples G5A, G5B, and G5C. The samples had been clarified by centrifugation at 3000 rpm before the measurements.

Absorbance measurements were made at room temperature (23 ± 2°) in a Cary 14 recording spectrophotometer which had been calibrated to ±2 Å in wavelength and better than ±0.5% in amplitude. Double-compartment cuvettes having optical paths of 5.00 and 10.0 mm in each compartment were used for difference absorbance measurements. Circular dichroism measurements were carried out in thermostated cells at temperatures known to ±0.5° in Cary 61 and Cary 6003 recording instruments which had been calibrated to ±2 Å in wavelength and to within 2% in amplitude with *d*-10-camphorsulfonic acid according to the method of Cassim and Yang (1969). Ellipticities are given as the measured quantities, θ, or as molar quantities per residue, either peptide or nucleotide, as $[\theta] = (100 \times \theta)/lc$, where l is the path length in cm, c is the molar concentration in residues, and the units are (deg cm^2)/dmol. Mean residue weights of 112 and 308 daltons were used for the protein and the DNA, respectively.

The fitting of measured CD spectra with reference spectra for contributing structural elements was done by computer with programs written in Fortran IV. In tests with synthetic data, the programs returned values of coefficients for the amounts of the various structures present which were uncertain to less than ±0.001%.

Results

Absorbance Spectrum of G5 Protein. A near-ultraviolet absorbance spectrum of G5 protein is presented in Figure 1, and specific absorbance coefficients for the maxima at 276 nm of three preparations are listed in Table I. These were obtained after corrections for light scattering had been made by plotting log *OD vs.* log λ for wavelengths longer than 320 nm and extrapolating into the absorbance range. Spectra of G5 protein samples measured in 0.8 M NaCl–buffer II or brought into other neutral aqueous buffers rapidly by Bio-Gel P4 chromatography, instead of the extensive dialysis necessary

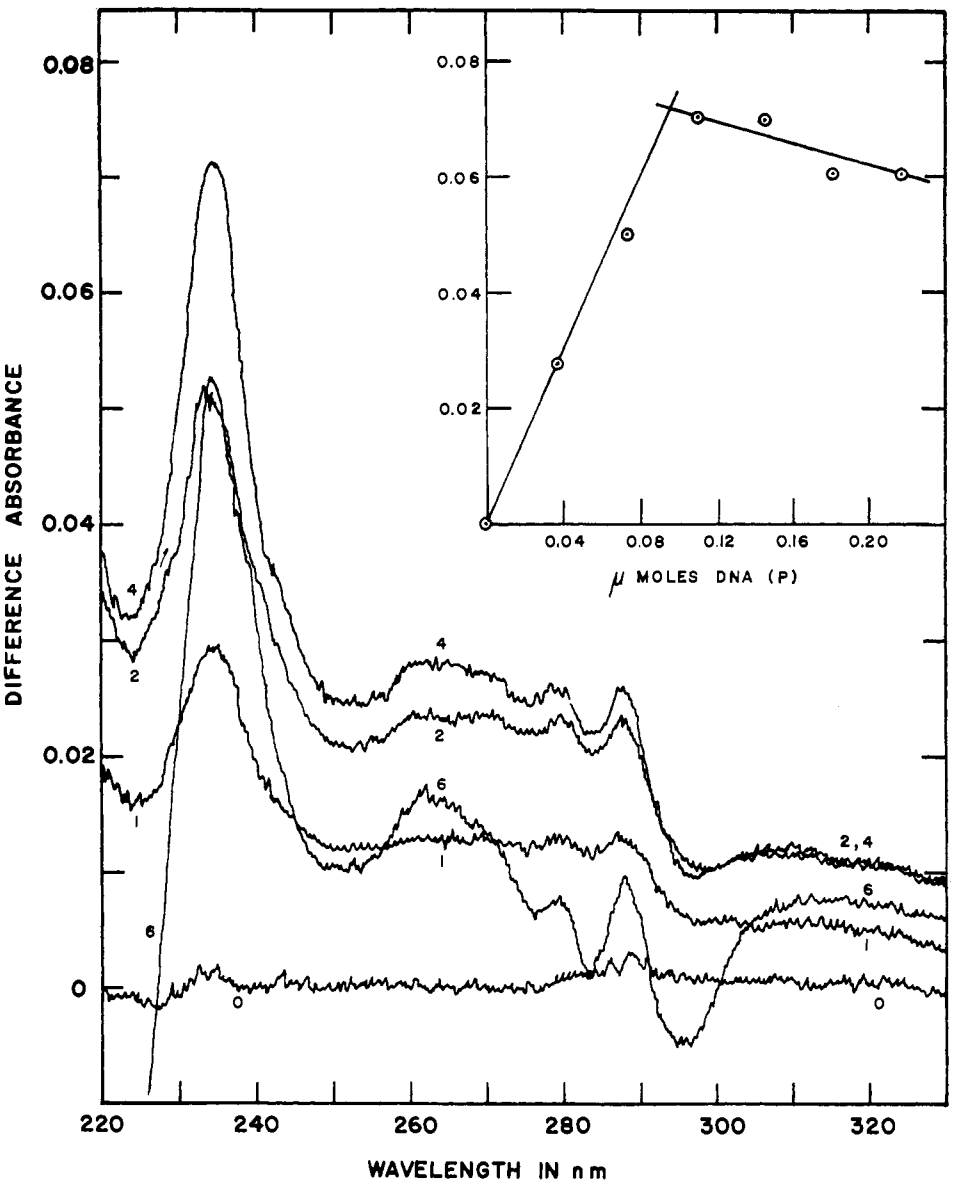

FIGURE 2: Difference spectral titration of 3.2 ml of 0.083 mg/ml of G5 protein at low ionic strength (2×10^{-4} M; buffer composition given in Table II) by successive 15-μl additions of 2.42×10^{-3} M (P) fd DNA. Numbers beside the scans indicate the number of additions of DNA solution. Scans 3 and 5, omitted from the figure for clarity, fell on scan 4 and slightly above scan 6, respectively. A description of the experiment is in the text. The insert shows the change at 235 nm as a function of DNA added during the titration after corrections for dilution had been made. The difference spectra become less positive on the addition of DNA beyond the equivalence point at low ionic strength; this has been observed repeatedly but not understood.

for the refractive index measurements, showed light-scattering contributions at 276 nm of less than 1%. The scattering is considered to come from a small fraction of denatured material and not necessarily from any specific protein–protein aggregation. The agreement between extinction coefficients corrected for scattering of different samples (Table I), and the value of $A^{1\%}_{1\,cm} = 7.3$ determined by biuret (Oey and Knippers, 1972; Alberts *et al.*, 1972), and a value of $A^{1\%}_{1\,cm} = 7.2$ one calculates from the amino acid composition (Oey and Knippers, 1972; Y. Nakashima and W. Konigsberg, personal communication; Gratzer, 1970), and the probable accuracy in the assumed refractive index increment are considered to reduce the uncertainty in spectrophotometrically determined G5 protein concentrations to about ±5%.

Changes in Absorbance on DNA–G5 Protein Binding. Two dual-compartment cells each containing G5 protein in one compartment and buffer in the other were placed in the sample and reference beams of a spectrophotometer, and the base line was recorded from 400 nm down to 225 nm. Equal amounts of DNA were added to the solution of G5 protein in the sample beam cell and to buffer in the reference beam cell. Each cell was gently rotated back and forth by hand about ten times and carefully replaced in the instrument, and the spectrum was scanned. Further additions of DNA gave titration difference spectra of the type shown in Figures 2 and 3 for experiments at low and high ionic strengths.

In all spectra, the DNA–protein mixtures showed greater extinction at wavelengths outside the absorbing region ($\lambda >$ 320 nm) than separated DNA and protein. The solutions of the mixtures remained clear visually, and plots of log *OD vs.* log λ from 320 to 360 nm had negative slopes of 4.1 ± 0.4 (dimensionless quantity) which is in accord with the inverse fourth power wavelength dependence of Rayleigh scattering and the wavelength dependence of dn/dc. The scattering was considered to derive from soluble protein–DNA complexes.

The large changes in the region below 320 nm have been

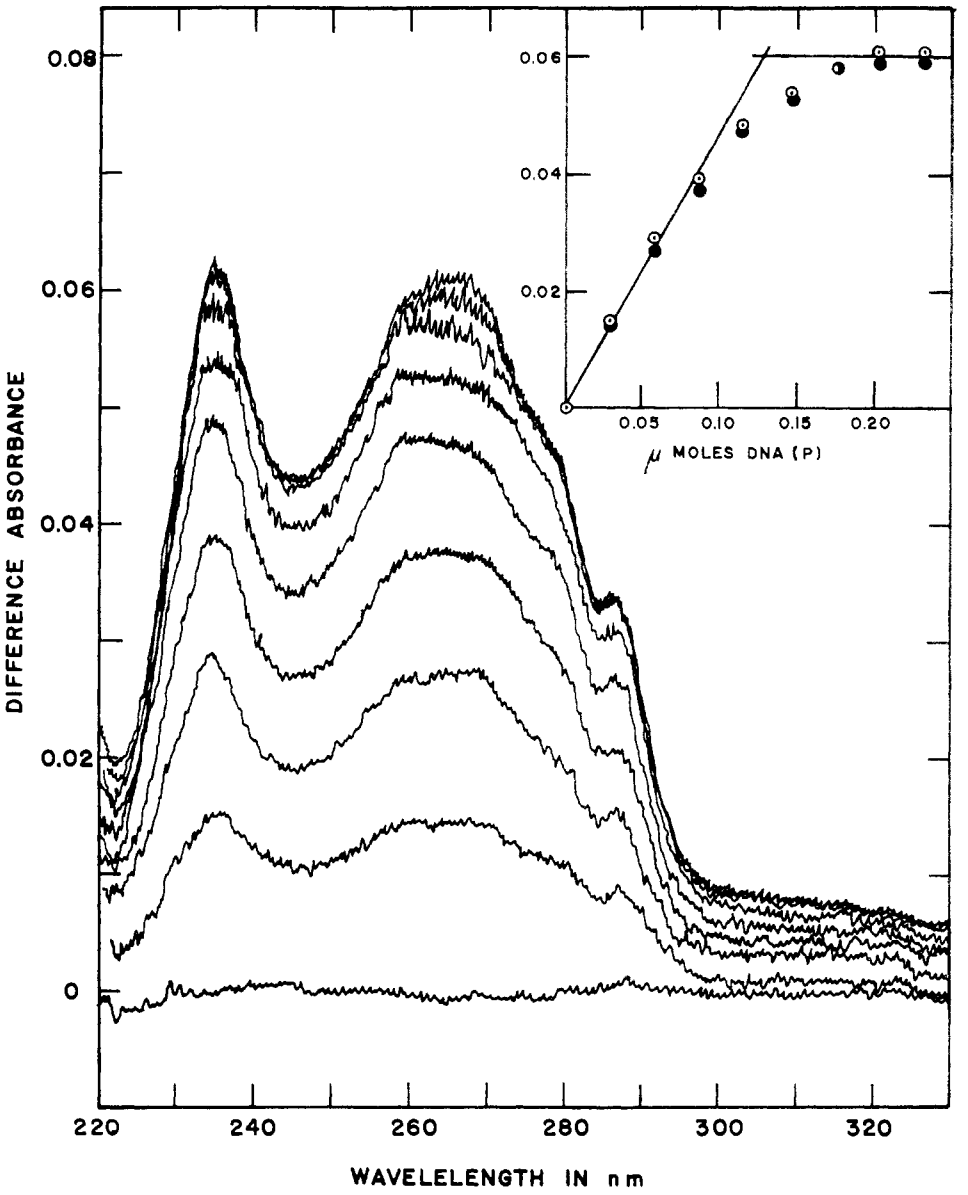

FIGURE 3: Difference spectral titration of 3.2 ml of 0.111 mg/ml of G5 protein at high ionic strength (0.26 M; buffer composition given Table II) by successive 12-μl additions of 2.42×10^{-3} M (P) fd DNA. Direct tracings of spectra after eight additions are presented; a ninth spectrum fell directly on the eighth. Similar spectra were obtained for titrations in the 0.2 M ionic strength buffer. The insert shows the changes in absorbance at 235 nm (○) and at 262 nm (●) as a function of DNA added.

plotted as a function of solution composition. The composition of the solution at the break point in the optical change has been taken as the composition of equivalent DNA–protein complexes. This involves the assumptions that all of the protein is native and that all regions of the circular single-stranded fd DNA are accessible. In this way difference absorbance titrations give a measure of the number of nucleotides bound per protein monomer of 10,000 daltons as summarized in Table V.

The difference spectra contain information on structural changes accompanying DNA–protein complex formation. First we present evidence for changes in tyrosine environment. The protein absorbance spectrum (Figure 1) has negative slopes with inflection points near 235, 279, and 287 nm, and positive slopes between 250 and 276 nm. The five tyrosyl groups are the collective source of these features, since tryptophan is not present and the four phenylalanyl groups make only minor contributions in this region (*cf.* Wetlaufer, 1962). If the spectrum were shifted 2 or 3 nm toward longer wavelengths, the difference between the shifted and the unshifted spectra would be positive or negative depending on whether the absorbance increases or decreases as one scans to shorter wavelengths, and would have local extrema where the absorbance spectrum is steepest. Such shifts toward longer wavelengths for model compounds containing protein chromophores have been observed with solvents which are less polar than water (Yanari and Bovey, 1960; Herskovits and Laskowski, 1962). Under low ionic strength conditions (Figure 2), a large positive maximum appears at 235 nm, and reproducible maxima at 279 and 288 nm are clearly evident. Under high ionic strength conditions (Figure 3), maxima are clearly seen at 235 and 289 nm, and a shoulder near 280 nm is apparent. These spectral features indicate that tyrosyls of G5 protein are transferred to a more nonpolar environment on formation of DNA–protein complexes.

The difference spectra also demonstrate that the DNA in the G5 protein–DNA complexes is highly hyperchromic. In the low ionic strength buffer, in which $\epsilon(P)_{260nm} = 8200$

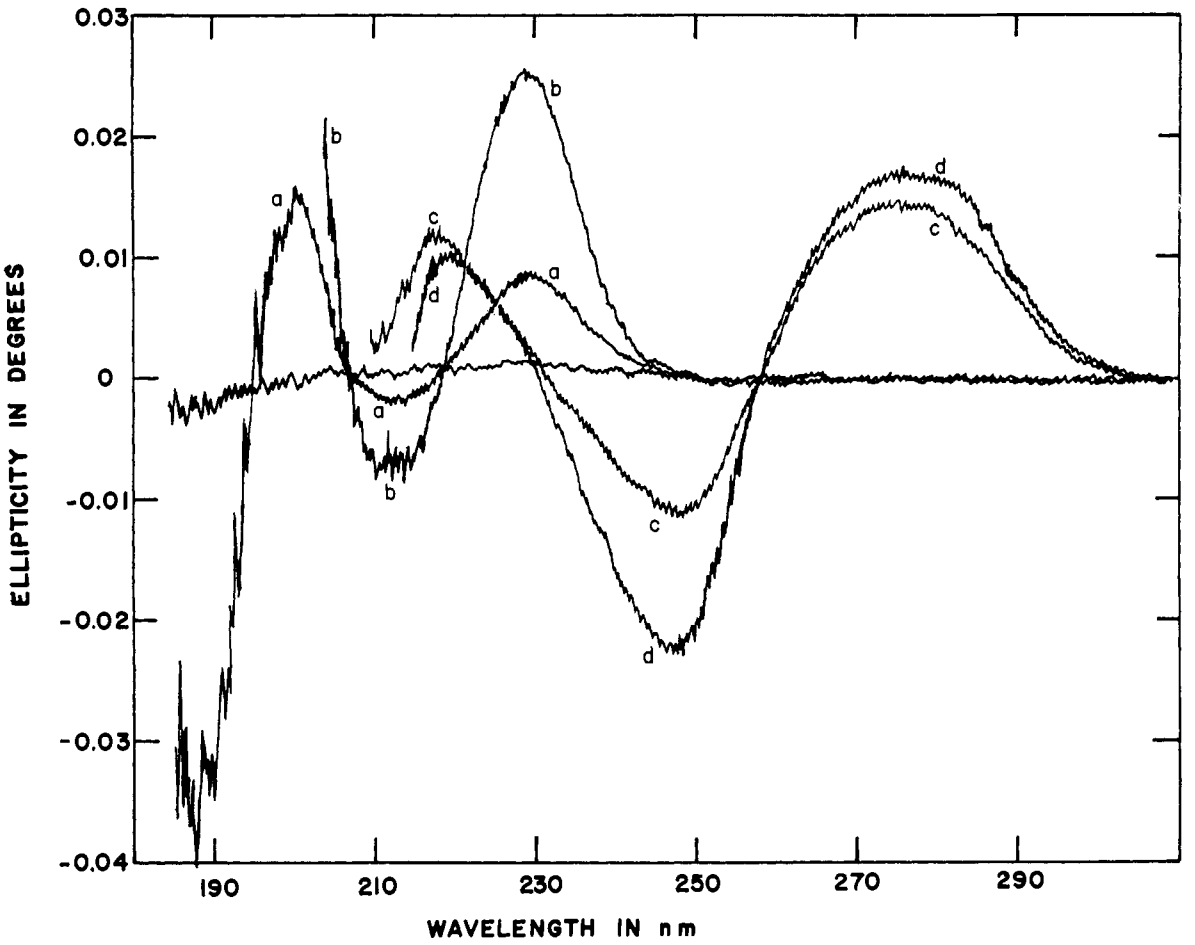

FIGURE 4: Circular dichroism in 0.15 M $NaClO_4$–0.02 M sodium phosphate (pH 7) of: (a) 0.23 mg/ml of G5 protein, 1-mm path; (b) 0.78 mg/ml of G5 protein, 1-mm path; (c) single-stranded fd DNA at a concentration of 0.91 OD_{260}/ml, 10-mm path; and (d) superhelical form I DNA from PM2 bacteriophage at 0.58 OD_{260}/ml, 10-mm path. Tracings of separate experiments done under the same instrumental conditions were made to form this figure.

for isolated single-stranded fd DNA, broad maxima are observed in the difference spectra near 260 nm, but after light-scattering corrections have been applied, there is little difference between the absorbance of the DNA–protein mixtures and the separated components near 260 nm (Figure 2). In high ionic strength buffers, however, the complexes absorb much more than the separated components in the 260-nm region. In the experiment shown in Figure 3, $\epsilon(P)_{260nm} = 6850$ for isolated fd DNA and a large positive maximum occurs at 262 nm. As indicated in the preceding paragraph, a shift of the protein spectrum to longer wavelengths would produce *negative* contributions at 260 nm, which would be small since the contribution of protein to the total absorbance at the equivalence point is only about 10%. On the basis of this and the fact that single-stranded fd DNA has a broad absorbance maximum at 260 nm which can undergo marked intensity changes, we ascribe the absorbance changes with a maximum at 262 nm to changes in DNA structure. *Minimum* values for the molar absorbance per nucleotide in the complexes can be obtained by computing $\epsilon(P)_{DNA}$ in complex $= \epsilon(P)_{260}(A_{260} + \Delta A_{262})/A_{260}$, where $\epsilon(P)_{260}$ is the molar nucleotide absorbance for isolated DNA in the given buffer, A_{260} is the absorbance for that concentration of free DNA equal to the concentration of bound DNA in the DNA–protein mixture, and ΔA_{262} is the difference absorbance after light-scattering corrections have been made. Such computations for titrations in three different solvents gave the results in Table II. The highest extinction reported for isolated single-strand fd DNA is $\epsilon(P) \simeq 8300$ (Hoffmann-Berling *et al.*, 1963) in 10^{-3} M NaCl, 22°, pH 7.

Circular Dichroism of G5 Protein. The CD spectrum of G5 protein is somewhat unusual in having a prominent positive band centered at 228 nm (Figure 4). Five spectra on four dif-

TABLE II: Hyperchromicity of fd DNA in Complexes with G5 Protein.

Ionic Strength of Buffer System[a]	$\epsilon(P)_{260}$[b] (Isolated fd DNA)	$(A_{260} + \Delta A_{262})/A_{260}$[c]	$\epsilon(P)$ (DNA in Complex)
2×10^{-4} M	8200	1.01 ± 0.01	8280
0.2 M	6950	1.15 ± 0.01	8000
0.26 M	6850	1.16 ± 0.02	7950

[a] The buffer systems designated by their ionic strengths are: (2×10^{-4} M) 1.5×10^{-4} M $NaClO_4$–1.5×10^{-5} M sodium phosphate (pH 7); (0.2 M) 0.15 M $NaClO_4$–0.02 M sodium phosphate (pH 7.0); (0.26 M) 0.23 M NaCl–0.01 M Tris-HCl–4.5 mM $MgCl_2$–0.5 mM EDTA–0.1 mM mercaptoethanol–1% glycerol (pH 7). [b] At 23°, based on literature cited in text. [c] Values calculated for each of the traces in the difference spectra and averaged. See text for definition of symbols. The errors indicated are standard deviations of six to nine traces.

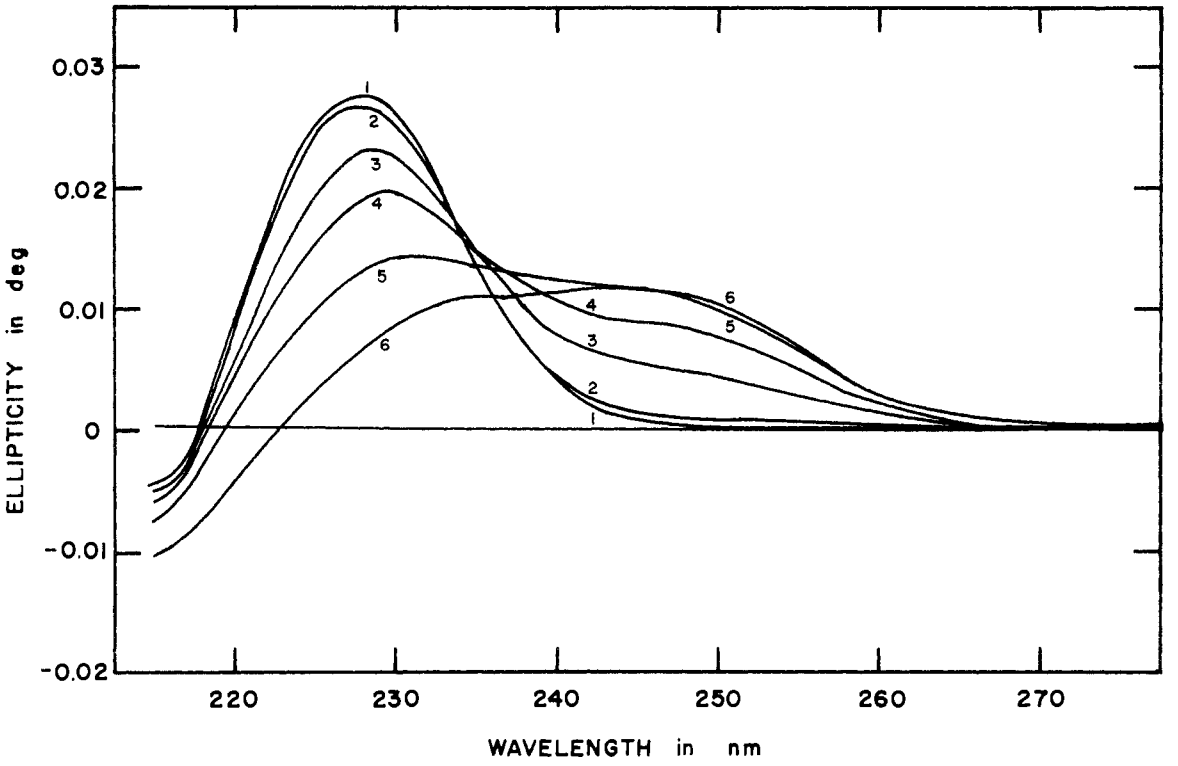

FIGURE 5: G5 protein CD as a function of pH in 0.15 M $NaClO_4$. 1.8-ml aliquots of 1.1 mg/ml of G5 protein in $NaClO_4$ were mixed with 0.2 ml of different concentrations of NaOH; the average of pH values measured immediately before and after the CD measurements were: 1, pH 8; 2, pH 8.9; 3, pH 10.0; 4, pH 10.6; 5, pH 11.2; and 6, pH 11.5. The light path was 10 mm.

ferent preparations of protein gave $[\theta]_{223} = 3362 \pm 150$ (deg cm^2)/dmol at neutral pH. The spectrum was found to be insensitive to changes in pH from 6 to 9.5 at 0.15 M ionic strength (H. T. Pretorius, unpublished observations), and to changes in ionic strength from 2×10^{-4} to 1 M at pH 7. However, the peak at 228 nm could be shifted to a broad positive band at 243 nm by raising the pH above 9.5 (Figure 5). This corresponds to a frequency shift of 2700 cm^{-1}, about the same as that observed for the $\pi-\pi^*$ transition at 222 nm of the phenolic group of tyrosine on deprotonation (*cf.* Donovan, 1969; Table III). Also, the long-wavelength maximum in the difference CD (~243 nm) is, within the experimental uncertainty, the same as that for the alkaline pH *vs.* neutral pH difference absorbance maximum of *N*-acetyltyrosine (Donovan, 1969; Figure 4). This correspondence makes it likely that phenolic transitions are the principle collective source of the 228-nm CD band. This assignment is strengthened in that positive CD bands near 228 nm have not been reported for histidine and cysteine, or for disulfide groups (Breslow, 1970; Beychok, 1965), and in that the G5 protein CD is unchanged over the pH range 6–9.5 which includes the normal p*K* values of histidine and cysteine. Positive CD bands in the region of 228 nm have been observed in many model compounds containing tyrosine, as can be seen from the summary of data in Table III, in which characteristic features of G5 protein CD have been included for comparison.

The pH dependence, the absence of other side chains which might contribute, and the extensive literature on tyrosine CD make the assignment of the 228-nm band to tyrosine sufficiently certain that it will be referred to below as the L_a tyrosine band, a $\pi-\pi^*$ transition of moderate intensity, to distinguish it from $\pi-\pi^*$ transitions of lower energy and intensity at 276 nm, the L_b band, and of higher energy and intensity near 200 nm, the B_a and B_b bands, according to the nomenclature of Platt (*cf.* Murrell, 1963, Chapter 6). It is nevertheless known that the $n-\pi^*$ transitions of the peptide chromophore (Holzwarth and Doty, 1965) and the L_a band of phenylalanine (Woody, 1972) are nearly degenerate with the tyrosyl L_a band, and these contribute to the CD in the 220–230-nm region.

An attempt has been made to estimate the overall conformation of G5 protein in terms of the fraction of residues, f_i, present in various reference structures. This was done by representing approximate CD amplitudes $[\theta_a(\lambda)]$, at different wavelengths, λ, as linear sums

$$[\theta_a(\lambda)] = \sum_{i=1}^{n} f_i[\theta_i(\lambda)]$$

where $[\theta_i(\lambda)]$ is the amplitude for the ith reference structure at wavelength λ, and then finding the best values of f_i so that the sum, over all wavelengths, of the squares of the differences between $[\theta_a(\lambda)]$ and the observed amplitudes $[\theta(\lambda)]$ is a minimum. The evaluation of the coefficients, f_i, was done by multiple linear regression analysis with the restriction that

$$\sum_{i=1}^{n} f_i = 1$$

but no restriction on the sign of the f_i. This type of approach to conformational analysis of proteins has been described by Magar (1968, 1971). Saxena and Wetlaufer (1971) solved the system of equations for $[\theta_i(\lambda)]$ after assigning the f_i for amounts of α-helix, β, and random structure present in three proteins of known crystal structure. We have used the reference spectra they obtained for these structures and presume that very similar results would have been obtained with the similar data of Greenfield and Fasman (1969) for poly-L-

TABLE III: CD of Tyrosine in Model Systems.[a,b]

Compound and Solvent	λ_p	$[\theta_p]$	λ_o	λ_t	$[\theta_t]$	λ_o	λ_p	$[\theta_p]$	λ_o	λ_t	$[\theta_t]$	Ref
N-Acetyltyrosinamide in water, pH 6	227	+11.7	–	215	+4.7							Simmons and Glazer (1967)
N-Acetyltyrosinamide in water, pH 7	225	+16.4	–	215	+7.9							Pflumm and Beychok (1969)
Cyclic (-Gly_5-L-Tyr-) in water	223	+6.1	–	–	–	210	–	–	–	200	−37.0	Ziegler and Bush (1971)
Linear Gly_5-L-Tyr-Gly_3 in water	226	+6.2	–	220	+3.3	–	210	+8.2	204	198	−36.6	Ziegler and Bush (1971)
$(Tyr)_n$ in trimethyl phosphate	231	+22.5	220	214	−4.2	211	202.5	+122.0	188.5			Damle (1970)
$(Tyr)_n$ in trimethyl phosphate	231	+29.5	220.5	214	−5.1	211	202.5	+150.0	188.5	185	−45.0	Engel *et al.* (1971)
$(Tyr)_n$ in methanol	230	+24.0	220.5	216	−4.0	211	203	+98.0	200			Shiraki and Imahori (1966)
$(Tyr)_n$ in water, pH 10.6, $\mu = 0.1$	230	+16.2	220	216	−0.7	214	200	+35.9				Friedman and Ts'o (1971a)
Theoretical for the RA α-helical conformation of $(Tyr)_n$ ($N = 10$) in water	231	+23.7	224	218	−16.5	214	201	+257.0	193	186	−233.0	Chen and Woody (1971) (Table VI and Figure 4)
G5 protein[c]	228	+3.4	219	212	−1.1	207	200	+6.6	193	188	−14.2	This study

[a] Data for long-wavelength bands near 276 nm have been omitted. λ_p, λ_o, and λ_t are wavelengths in nm for peaks, crossover points, and troughs, respectively. $[\theta_p]$ and $[\theta_t]$ are the molar ellipticities per tyrosyl residue $\times 10^{-3}$ in (deg cm^2)/dmol. [b] A dash (–) indicates not applicable, whereas a blank indicates data not available. [c] Data for G5 protein included for comparison of major features. Ellipticities here expressed per peptide bond.

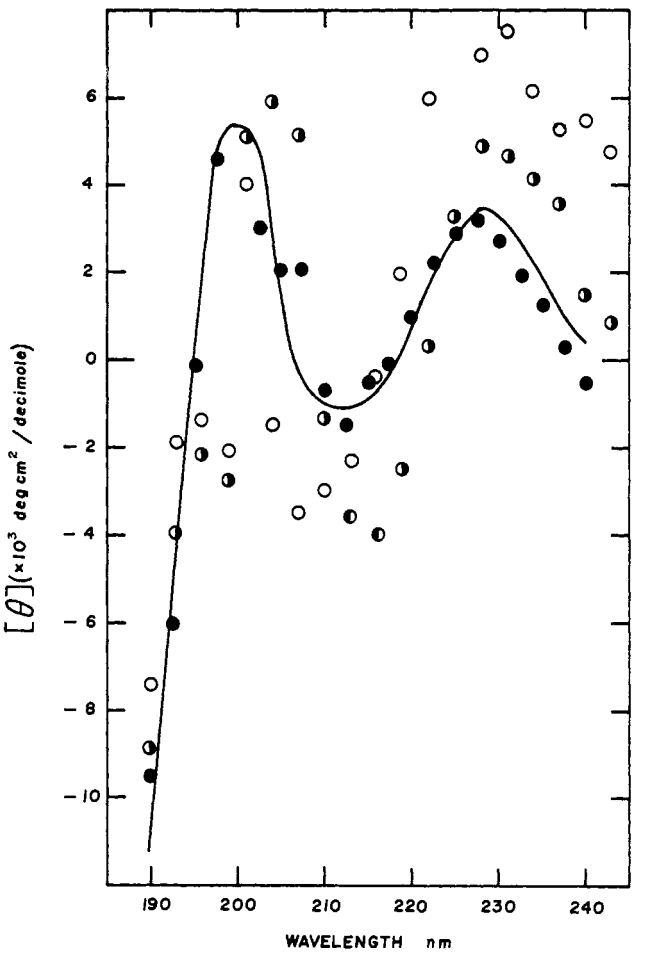

FIGURE 6: Comparison of the measured CD spectrum for G5 protein (solid curve) with some of the spectra computed as described in the text. The symbols ◑, ○, and ● correspond to computations 4, 5, and 10 listed in Table IV.

lysine. We have also used the reference data of Chen *et al.* (1972) who used five proteins of known crystal structure to obtain CD reference spectra which have significant quantitative differences from those of Saxena and Wetlaufer.

The results of ten computations are presented in Table IV. The first computation with the data of Chen *et al.* (1972) gave a small but negative coefficient for the amount of α-helix present. Since the reference used was a right-handed α helix, the computation was repeated with signs of the reference amplitudes reversed with the assumption that, to the first approximation, the CD of a left-handed α helix will be the mirror image of a right-handed one. The effect of the contributions from tyrosyl groups was then tested by using, as an additional reference, a CD spectrum for polytyrosine (computations 3 and 4). The polytyrosine spectrum was chosen because it was available over the wavelengths needed and because it looks remarkably like the G5 protein CD spectrum (Table III). Four parameter fits (computations 3 and 4) and a three parameter fit (computation 5) with α helix omitted showed reduced standard errors in the computed coefficients. The computations were repeated with the data of Saxena and Wetlaufer (computations 6–10) with better fits to the experimental spectrum for G5 protein but about the same apparent amounts of different structures present. An idea of the quality of some of the fits can be obtained from Figure 6. The major result of the analysis is that the CD of G5 protein indicates the presence of little α helix of either sense (Table IV).

TABLE IV: Fitting of G5 Protein CD with Reference Spectra.[a]

Computation No.	f α Helix[b]	f β Structure	f Random Structure	f Tyr-like Structure[c]	Ref for α, β, and Random
1	-0.03 ± 0.04 (R)	0.65 ± 0.23	0.38 ± 0.20	--	*d*
2	0.06 ± 0.04 (L)	0.72 ± 0.20	0.22 ± 0.2	--	*d*
3	-0.04 ± 0.02 (R)	0.47 ± 0.14	0.45 ± 0.12	0.11 ± 0.02	*d*
4	0.05 ± 0.03 (L)	0.49 ± 0.13	0.35 ± 0.14	0.11 ± 0.02	*d*
5	--	0.33 ± 0.12	0.56 ± 0.11	0.11 ± 0.02	*d*
6	-0.15 ± 0.04 (R)	0.59 ± 0.06	0.44 ± 0.1	--	*e*
7	0.17 ± 0.03 (L)	0.48 ± 0.03	0.35 ± 0.05	--	*e*
8	-0.06 ± 0.02 (R)	0.40 ± 0.04	0.61 ± 0.02	0.06 ± 0.01	*e*
9	0.05 ± 0.03 (L)	0.34 ± 0.03	0.55 ± 0.06	0.06 ± 0.01	*e*
10	--	0.29 ± 0.01	0.64 ± 0.01	0.07 ± 0.01	*e*

[a] Computations made over the range 190–243 nm at 3-nm intervals with published data of Chen *et al.* (1972) and at 2.5-nm intervals with digital data kindly supplied by V. P. Saxena. The uncertainties are the standard errors in the coefficient for a given computation. [b] There is no restriction on the sign of coefficients determined by least-square fitting. A negative sign for the amount of right-handed helix was taken, at least for the purpose of calculations, as indicating the presence of left-handed helix. Computations 2 and 4 were repeats of 1 and 3 with signs for the helix reference CD reversed. [c] Data of Shiraki and Imahori (1967) for polytyrosine at pH 10.6, $\mu = 0.1$. [d] Chen *et al.* (1972) and Table V. [e] Saxena and Wetlaufer (1971) and Figure 1A.

CD Changes on Mixing DNA with G5 Protein. CD spectra for isolated single-stranded fd DNA and PM2 DNA are also presented in Figure 4. Single-stranded fd DNA at moderate ionic strength has a bimodal (Tinoco *et al.*, 1963; Warshaw *et al.*, 1965), conservative (Bush and Brahms, 1967) CD exciton band centered near 260 nm similar to double-stranded PM2 DNA and other DNAs (Brahms and Mommaerts, 1964). The protein spectrum is very weak in this near-ultraviolet region. There are crossovers in the DNA CD spectra where the tyrosyl L_a band of the protein has its maximum.

In Figure 7 a CD titration is shown in which fd DNA was added to G5 protein. The drop in ellipticity at 228, 230, and 232 nm was plotted as a function of solution composition and the apparent equivalence point was the same for the plots. The largest change was found at 230 nm where corrections for the CD contributions of DNA, if free, would be 3% of the L_a band of isolated protein at the highest DNA concentrations reached. The reduction in ellipticity caused by the DNA was about 40% of the total peak amplitude. The original amplitude could be recovered by adding solid NaCl to a concentration of 0.7 M or greater to the DNA–protein mixture. The results of CD titrations are summarized in Table V, together with results from the difference absorbance titrations.

Even though these changes in protein CD near 230 nm are prominent and easily measured, they are quite small per peptide residue ($\Delta[\theta]_{230} \sim 1200$ (deg cm^2)/dmol). Aside from these small changes, amplitudes for DNA–protein mixtures in the region used for protein conformational analysis above (190–243 nm) show only minor deviations from additivity of the CD for separated components. It is concluded that the overall secondary conformation of the protein does not change significantly on complex formation.

On the other hand, protein–DNA binding can cause changes in CD bands for DNA in accord with the hyperchromicity changes. The G5 protein, which has essentially no CD near 260 nm, substantially reduces the CD exciton band of single-stranded fd DNA. The effect is demonstrated in Figure 8 where the spectra of three protein–DNA mixtures have been plotted in terms of a constant amount of DNA. This reduction would be expected if the bases become unstacked during the binding.

PM2 DNA is a super twisted circular duplex DNA of 6×10^6 daltons (Espejo and Canello, 1968) containing no single-strand breaks in either of the strands. It was chosen as a convenient double-stranded DNA molecule for an experiment to see whether any of the spectral changes described above for mixing single-stranded DNA and G5 protein would occur with native unmelted double-stranded DNA. A CD spectrum of a mixture containing 0.23 mg/ml of G5 protein and 0.02 mg/ml of PM2 DNA in neutral aqueous buffer at room temperature could not be distinguished from that computed from the spectra of isolated components.

TABLE V: Summary of CD and Difference Absorbance (DA) Titrations.

Sample	Ionic Strength of Buffer (M)[a]	Protein Concn (mg/ml)	Method	DNA(P): Protein[b]
G5D	0.20	0.210	CD	4.21
G5F	0.08	0.045	CD	3.95
G5F	0.08	0.095	CD	3.91
G5G	0.04	0.098	CD	3.95
G5G	2×10^{-4}	0.083	CD	4.07
G5G	2×10^{-4}	0.083	DA	3.8
G5G	0.20	0.092	DA	3.9
G5G	0.26	0.111	DA	3.8
			Av[c]	3.95 ± 0.14

[a] Composition of solvent given in Table II, except for 0.08 and 0.04 M which were 10- and 20-fold dilutions of buffer II containing 0.8 M NaCl. [b] From plots of the types shown in the inserts of Figures 2, 3, and 5. [c] Standard deviation given for equal weighting of all eight values.

Discussion

Stoichiometry. The uncertainties in protein and DNA concentrations and in the definition of apparent equivalence points by spectral titration give an overall uncertainty of about 10% in the ratio of DNA to protein in the complex. After our computations based on 10,000 daltons/protein monomer

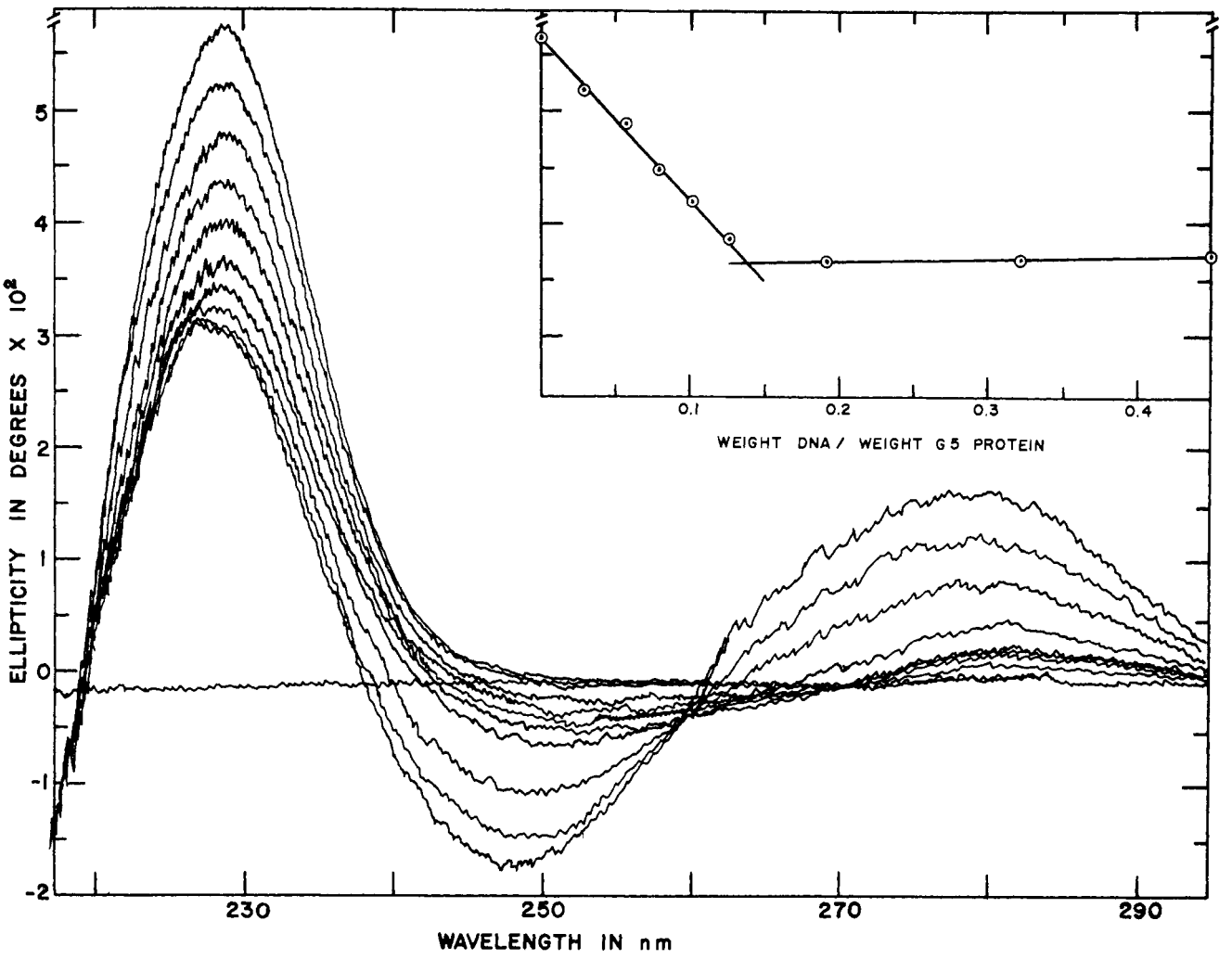

FIGURE 7: CD titration of 1.56 ml of a solution containing 0.21 mg/ml of G5 protein by additions of a solution of 0.27 mg/ml of fd DNA. Light path length was 10 mm. There were five increments of 20 μl, one of 50 μl, and three of 100 μl. Spectra were recorded immediately after the addition of DNA and gentle mixing. The initial scan shows the highest amplitude at 228 nm and the final scan, the lowest. The insert gives the titration plot of the same data at 230 nm after corrections for dilution had been made. The ordinate scale is for both the spectra and the insert.

were complete, data on the amino acid sequence of G5 protein became available to us (Y. Nakashima and W. Konigsberg, private communication) which show G5 protein to contain 88 residues with a total weight of 9830 daltons. This increases our average ratio to 4.02 ± 10% nucleotides per monomer of G5 protein. Various ways of handling the protein did not lead to an increased nucleotide:protein ratio. The ratio is in agreement with values reported by Oey and Knippers (1972) and by Alberts *et al.* (1972) whose methods, although different, are subject to the same assumptions regarding nativeness of protein and complete accessibility of nucleotides. Accurate knowledge of the ratio is important for considerations of symmetry in the complexes. The micrographs of Delius (Alberts *et al.*, 1972) show that these complexes are filaments with twofold symmetry, which makes a decision between an odd or even number of nucleotide binding sites important and which prompted extra effort on this point in the present study. Some of the consequences of four sites per monomer on the structure of the nucleoprotein have been discussed by Alberts *et al.* (1972).

Conformation of Gene 5 Protein. A prominent feature of the CD of G5 protein is the positive band at 228 nm, which has been assigned as a tyrosyl L_a band. Since *N*-acetyltyrosinamide itself shows such strong optical activity, about half that calculated for poly-L-tyrosine (*cf.* Table III, above; Simmons and Glaser, 1967; Pflumm and Beychok, 1969), and isolated tyrosyls in oligopeptides show similar optical activity (Ziegler and Bush, 1971), it seems likely that much of the optical activity of the tyrosyl L_a band derives from interactions between a given phenolic chromophore and its local amides. The five tyrosyls of G5 protein are separated along the sequence appearing at positions 26, 34, 41, 56, and 61 (Y. Nakashima and W. Konigsberg, personal communication) so that interactions between the phenolic chromophores would only come about by intrachain folding or interchain associations.

Conformation assignments of peptide chains from CD are hazardous, principally because of contributions from aromatic chromophores. Nevertheless useful characterizations can be made by relating the protein CD to probable contributions of frequently encountered reference structures. We have used reference CD spectra for α-helix, β, and random conformations and reference CD for tyrosyl contributions for computer fitting of the G5 protein CD. The results in Table IV show that, according to this analysis of its CD, G5 protein may have anywhere from 30 to 70% of its residues in the β conformation and anywhere from 20 to 65% in ran-

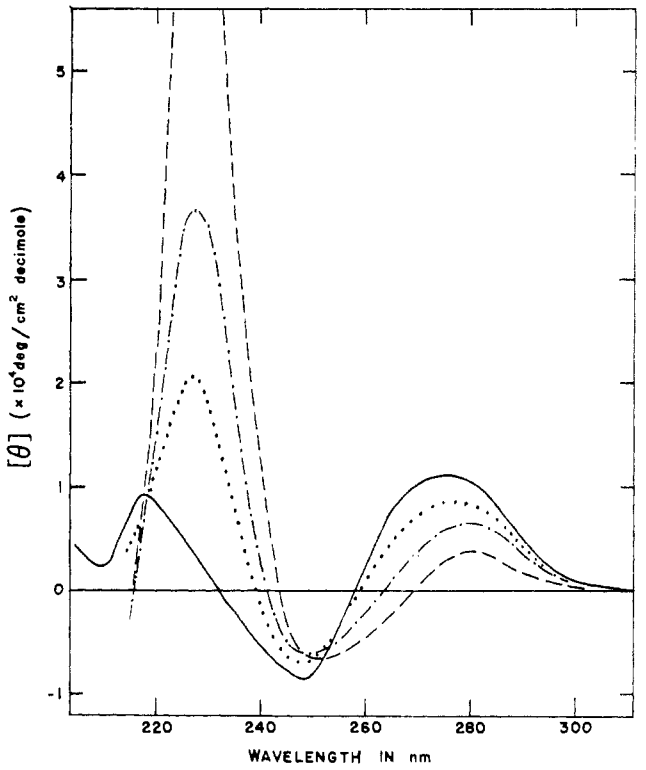

FIGURE 8: Demonstration of the effect of G5 protein on the near-uv exciton band of isolated single-stranded fd DNA in 0.15 M $NaClO_4$–0.02 M sodium phosphate (pH 7) were replotted in terms of a constant concentration of nucleotide residues. The solid curve is for the isolated DNA and the others are for nucleotide:protein monomer molar ratios of 9.4 (*****), 4.8 (-*-*-), and 3.7 (- - -).

dom conformations. The low fraction of residues present in α helices, at the most 20% and probably less than 10% according to the CD analysis, is consistent with the sequence data, cited above, which show six proline residues located at positions 8, 25, 42, 54, 58, and 86 along the chain.

Conformation Changes on DNA–Protein Binding. The spectral results indicate that only small changes occur in overall protein structure on binding. The reduction in the 228-nm protein band probably results from a change in local environment or orientation of the tyrosyl chromophores, either through direct interaction with DNA or through secondary effects. Since the complexes formed appear filamentous by electron microscopy (Alberts *et al.*, 1972), it is not expected that absorbance flattening or differential scattering of left and right circularly polarized light (Duysens, 1956; Urry and Ji, 1968; Gordon and Holzwarth, 1971; Day and Hoppensteadt, 1972) would contribute significantly to the reduction. The fine structure in the difference absorbance spectra are examples of a change to a less aqueous solvent environment for tyrosine on complex formation. Solvent perturbation studies (Herskovits and Laskowski, 1962) are needed to establish the change in solvent environment more quantitatively.

The hyperchromic shift near 260 nm and the large CD changes in this region on DNA–protein binding are considered to be the result of reduced base–base interactions concomitant with the unstacking of the bases in single-stranded fd DNA. For reasons given in the Results section, the molar extinction per nucleotide in the G5 protein–fd DNA complex is considered to be at least 8200 l./(mol cm), or as high as that observed for hyperchromic denatured single-stranded fd DNA in low ionic strength aqueous solvents (Hoffmann-Berling *et al.*, 1963; Day, 1969; Schaller *et al.*, 1969). Oey and Knippers (1972) and Alberts *et al.* (1972) have also observed hyperchromic shifts on G5 protein–DNA complex formation.

Involvement of Tyrosine in DNA–Protein Binding. Recent model studies by Friedman and Ts'o (1971b), who showed that poly-L-tyrosine forms complexes with denatured DNA but not with native double-stranded DNA, by Hélène (1971), who demonstrated by fluorescence methods that aromatic amino acid side chains can bind to DNA, and by Adler *et al.* (1972), who proposed that tyrosine may be involved in the DNA binding site of the *lac* repressor, suggest that tyrosyl residues might frequently be involved in the binding of nucleic acids with various proteins. The results of the present study certainly indicate that tyrosine might be critically involved in DNA–protein binding. In addition, it is known that basic groups in proteins are involved in nucleic acid binding in many systems. For the G5 protein–DNA system, the electrostatic nature of the binding is in fact used in the preparation of G5 protein by salt elution from DNA columns. It may be that many recognition and interaction sites of proteins and nucleic acids involve tyrosyl groups as well as the basic groups necessary for charge neutralization.

Comparison with the fd Virion and the Role of G5 Protein in Bacteriophage Morphogenesis. The structure of the G5 protein–fd DNA complex is quite different from the protein–DNA complex which constitutes the virus itself. The single-stranded DNA in the filamentous virion has been shown to be strongly hypochromic (Day, 1969) whereas in the G5 protein complex it is strongly hyperchromic. G5 protein–fd DNA complexes appear as thin filaments in electron micrographs with contour lengths 20% greater (Alberts *et al.*, 1972) than those of the virion (Frank and Day, 1970). The spectral and micrographic data indicate that DNA in the G5 protein complex is unstacked and stretched along the filament axis, whereas the DNA in the virion is base stacked and has a short base–base separation along the virion filament axis. The protein conformations in the two complexes are also different. Whereas the G5 protein, according to the analysis above, contains little α helix, the major coat protein in the virion is highly α helical (Marvin, 1966; Day, 1966, 1969; Asbeck *et al.*, 1969).

It seems reasonable to consider that G5 protein–fd DNA complexes exist *in vivo* which are similar to the complexes formed *in vitro*, although this must be verified. From the available results, including those cited in the introduction and those of Smilowitz *et al.* (1972), who have shown that the major viral coat protein is associated with the inner membrane of the infected cell, one might envisage viral assembly as involving the conversion of unstacked single-helical DNA maintained by non-α-helical protein subunits, coded by one viral gene, into stacked single-helical DNA maintained by highly α-helical protein subunits coded by another gene. The product of a third gene, which constitutes a minor protein of the viral coat, may also have a critical role in the assembly reaction (Rossomando and Zinder, 1968). The final product, a progeny virion, would be extruded into the medium during the course of such a subunit exchange reaction.

Added in Proof

The recent results of D. Pratt, P. Laws, and J. Griffith (1973) provide good evidence for the presence of complexes between gene 5 protein and viral DNA in the infected bacterium.

Acknowledgments

It is a pleasure to acknowledge the excellent assistance of Margarete Klein during the course of this study. For providing materials, suggestions, and manuscripts for publication, I am very grateful to Bruce M. Alberts, A. Keith Dunker, Rolf Knippers, W. Konigsberg, D. A. Marvin, H. T. Pretorius, and John S. Salstrom.

References

Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M., and Schmitz, A. (1972), *Nature (London) 237*, 322.

Alberts, B. M., Frey, L., and Delius, H. (1972), *J. Mol. Biol. 68*, 139.

Alberts, B. M., and Herrick, G. (1970), *Methods Enzymol. 21*, 198.

Asbeck, F., Beyreuther, K., Köhler, H., von Wettstein, G., and Braunitzer, G. (1969), *Hoppe-Seyler's Z. Physiol. Chem. 350*, 1047.

Beaudoin, J. (1970), Ph.D. Thesis, University of Wisconsin.

Beychok, S. (1965), *Proc. Nat. Acad. Sci. U. S. 53*, 999.

Brahms, J., and Mommaerts, W. F. H. M. (1964), *J. Mol. Biol. 10*, 73.

Breslow, E. (1970), *Proc. Nat. Acad. Sci. U. S. 67*, 493.

Bush, C. A., and Brahms, V. (1967), *J. Chem. Phys. 46*, 79.

Cassim, J. Y., and Yang, J. T. (1969), *Biochemistry 8*, 1947.

Chen, A. K., and Woody, R. W. (1971), *J. Amer. Chem. Soc. 93*, 29.

Chen, Y. H., Yang, J. T., and Martinez, H. M. (1972), *Biochemistry 11*, 4120.

Damle, V. N. (1970), *Bioploymers 9*, 937.

Day, L. A. (1966), *J. Mol. Biol. 15*, 395.

Day, L. A. (1969), *J. Mol. Biol. 29*, 265.

Day, L. A., and Hoppensteadt, F. A. (1972), *Biopolymers 11*, 2131.

Donovan, J. W. (1969), *in* Physical Principles and Techniques of Protein Chemistry, Part A, Leach, S. J., Ed., New York, N. Y., Academic p 101.

Duysens, L. N. M. (1956), *Biochim. Biophys. Acta 19*, 1.

Engel, J., Leihl, E., and Sorg, C. (1971), *Eur. J. Biochem. 21*, 22.

Espejo, R. T., and Canello, E. S. (1968), *Virology 34*, 738.

Frank, H., and Day, L. A. (1970), *Virology 42*, 144.

Friedman, S., and Ts'o, P. O. P. (1971a), *Biochem. Biophys. Res. Commun. 42*, 510.

Friedman, S., and Ts'o, P. O. P. (1971b), *Biochemistry 10*, 3099.

Gierer, A., and Schramm, G. (1956), *Z. Naturforschg. B 11*, 138.

Gordon, J., and Holzwarth, G. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 2365.

Gratzer, W. B. (1970), *in* Handbook of Biochemistry, Sober, H. A., Ed., 2nd ed, Cleveland, Ohio, Chemical Rubber Co., p B74.

Greenfield, N., and Fasman, G. D. (1969), *Biochemistry 8*, 4108.

Hélène, C. (1971), *Nature (London), New Biol. 234*, 120.

Henry, T. J., and Pratt, D. (1969), *Proc. Nat. Acad. Sci. U. S. 62*, 800.

Herskovits, T. T., and Laskowski, M. (1962), *J. Biol. Chem. 237*, 2481.

Hoffmann-Berling, H., Marvin, D. A., and Dürwald, H. (1963), *Z. Naturforschg. B 18*, 876.

Holzwarth, G., and Doty, P. (1965), *J. Amer. Chem. Soc. 87*, 218.

Magar, M. E. (1968), *Biochemistry 7*, 617.

Magar, M. E. (1971), *J. Theor. Biol. 35*, 105.

Marvin, D. A. (1966), *J. Mol. Biol. 15*, 8.

Murrell, J. N. (1963), The Theory of the Electronic Spectra of Organic Molecules, London, Methuen and Co. Ltd., Chapter 6.

Oey, J. L., and Knippers, R. (1972), *J. Mol. Biol. 68*, 125.

Pflumm, M. N., and Beychok, S. (1969), *J. Biol. Chem. 244*, 3973.

Pratt, D., and Erdahl, M. S. (1968), *J. Mol. Biol. 37*, 181.

Pratt, D., Laws, P., and Griffith, J. (1973), *J. Mol. Biol.* (in press).

Pratt, D., Tzagoloff, H., and Beaudoin, J. (1969), *Virology 39*, 42.

Rossomando, E. F., and Zinder, N. D. (1968), *J. Mol. Biol. 36*, 387.

Salstrom, J. S., and Pratt, D. (1971), *J. Mol. Biol. 61*, 489.

Saxena, V. P., and Wetlaufer, D. B. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 969.

Schaller, H., Voss, H., and Gucker, S. (1969), *J. Mol. Biol. 44*, 445.

Shiraki, M., and Imahori, K. (1966), *Sci. Pap. Coll. Gen. Educ., Univ. Tokyo 16*, 215 (as cited by Chen and Woody, 1971).

Simmons, N. S., and Glazer, A. N. (1967), *J. Amer. Chem. Soc. 89*, 5040.

Smilowitz, H., Carson, J., and Robbins, P. W. (1972), *J. Supramol. Struct. 1*, 8.

Staudenbauer, W. J., and Hofschneider, P. H. (1973), *Eur. J. Biochem. 34*, 569.

Timasheff, S. N. (1970), *in* Handbook of Biochemistry, Sober, H. A., Ed., 2nd ed, Cleveland, Ohio, Chemical Rubber Co., p C67.

Tinoco, I., Woody, R. W., and Bradley, D. G. (1963), *J. Chem. Phys. 38*, 1317.

Urry, D. W., and Ji, T. H. (1968), *Arch. Biochem. Biophys. 128*, 802.

Warshaw, M. W., Bush, C. A., and Tinoco, I. (1965), *Biochem. Biophys. Res. Commun. 18*, 633.

Wetlaufer, D. B. (1962), *Advan. Protein Chem. 17*, 304.

Woody, R. W. (1972), *Biopolymers 11*, 1149.

Yanari, S., and Bovey, F. A. (1960), *J. Biol. Chem. 10*, 2818.

Zeigler, S. M., and Bush, C. A. (1971), *Biochemistry 10*, 1330.