The Binding Affinity of Ff Gene 5 Protein Depends on the Nearest-Neighbor Composition of the ssDNA Substrate

Tung-Chung Mou, Carla W. Gray, and Donald M. Gray*
Department of Molecular and Cell Biology, The University of Texas at Dallas, Richardson, Texas 75083-0688 USA

ABSTRACT The Ff gene 5 protein (g5p) is considered to be a nonspecific single-stranded DNA binding protein, because it binds cooperatively to and saturates the Ff bacteriophage single-stranded DNA genome and other single-stranded polynucleotides. However, the binding affinity $K\omega$ (the intrinsic binding constant times a cooperativity factor) differs by over an order of magnitude for binding to single-stranded polynucleotides such as poly[d(A)] and poly[d(C)]. A polynucleotide that is more stacked, like poly[d(A)], binds more weakly than one that is less stacked, like poly[d(C)]. To test the hypothesis that DNA base stacking, a nearest-neighbor property, is involved in the binding affinity of the Ff g5p for different DNA sequences, $K\omega$ values were determined as a function of NaCl concentration for binding to six synthetic sequences 48 nucleotides in length: dA_{48}, dC_{48}, $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$. The binding affinities of the protein for these sequences were indeed found to be related to the nearest-neighbor compositions of the sequences, rather than to simple base compositions. That is, the g5p binding site, which is spanned by four nucleotides, discriminates among these sequences on the basis of the relative numbers of nearest neighbors (AA, CC, and AC plus CA) in the sequence. The results support the hypothesis that the extent of base stacking/unstacking of the free, nonbound ssDNA plays an important role in the binding affinity of the Ff gene 5 protein.

INTRODUCTION

The multifunctional Ff gene 5 protein (g5p) is produced by the closely related M13, F1, and fd strains of filamentous bacterial viruses, collectively known as Ff viruses. The g5p is among the smallest of proteins that bind nonspecifically, cooperatively, and preferentially to single-stranded DNA (ssDNA) and are required in the life cycles of viruses, prokaryotes, and eukaryotes (Alberts et al., 1972; Wold and Kelly, 1988; Heyer and Kolodner, 1989; Heyer et al., 1990). The g5p monomer has 87 amino acids and a molecular weight of 9690, and it exists as a stable dimer, even at concentrations as low as 5×10^{-10} M (Terwilliger, 1996). It is a key regulator of the expression and replication of the Ff phage genome during infection of *Escherichia coli* bacteria. The protein has been most studied for its non-sequence-specific binding ability and high cooperativity, which results in the contiguous binding of g5p dimers to coat the viral ssDNA genome. The resulting superhelical structure, a precursor to the mature phage particle, has been found by electron microscopy to be a left-handed helix (Gray, 1989), it has been studied by solution scattering techniques (Gray et al., 1982; Olah et al., 1995), and it has been modeled by several groups (Skinner et al., 1994; Folmer et al., 1994; Guan et al., 1995; Olah et al., 1995). Moreover, the g5p recognizes specific DNA and RNA sequences, as discussed below.

Received for publication 19 June 1998 and in final form 18 November 1998.

Address reprint requests to Dr. Donald M. Gray, Department of Molecular and Cell Biology, University of Texas at Dallas, F031, P.O. Box 830688, Richardson, TX 75083-0688. Tel.: 972-883-2513; Fax: 972-883-2409; E-mail: dongray@utdallas.edu.

0006-3495/99/03/1537/15 $2.00

High-resolution crystallographic data are available for the Ff g5p and mutants in the absence of nucleic acid (Skinner et al., 1994; Guan et al., 1994; Zhang et al., 1996). One feature that is shared with known structures of other single-stranded nucleic acid-binding proteins (Pf3 ssDBP, T4 g32p, adenovirus ssDBP, RNA-binding domains of the small nuclear ribonucleoprotein U1A, the ssDNA-binding domain of *E. coli* topoisomerase I, the *E. coli* SSB protein, human replication protein A, and *E. coli* rho factor) is that they all use aromatic side chains from proximal β-strands to interact with bases or to form base-binding concavities (Folmer et al., 1995, 1997; King and Coleman, 1988; Shamoo et al., 1995; Yu et al., 1995; Raghunathan et al., 1997; Bochkarev et al., 1997; Briercheck et al., 1998; Allison et al., 1998).

G5p binds to specific RNA sequences to autoregulate translation of its own mRNA, and it regulates translation of mRNAs for other viral genes (Zaman et al., 1991, 1992). The RNA binding sequences have weak homology and are devoid of obvious secondary structure, but are rich in uracils. The translational regulation by g5p is very different from that described for the autoregulation of phage T4 g32p, which depends on binding to an unstructured sequence of unique length terminated by hairpins (von Hippel et al., 1982). In vitro, the g5p binds preferentially to the 16-mer DNA analog of the gene 2 mRNA binding sequence, so this DNA sequence might function as a nucleation site for cooperative binding of g5p to nascent viral ssDNA molecules (Michel and Zinder, 1989). A (G + C)-rich dsDNA hairpin of 32 bp is oriented at one end of the g5p · ssDNA intracellular complex (Bauer and Smith, 1988) and remains at one end of the mature phage after exchange of g5p for the g8p coat protein (Ikoku and Hearst, 1981; Webster et al., 1981). An analysis of the effects of g5p mutations has

identified a set of five amino acid residues that differentially affect the processes of DNA packaging and RNA translational repression (Stassen et al., 1992). These residues are located in a region of the protein distant from the DNA binding site, but they apparently affect the ability of the protein to conform to different nucleic acid structures and/or sequences.

How the protein distinguishes among different DNA or RNA sequences is not known. Like most other single-strand DNA-binding proteins, g5p binds more tightly to some homopolymers (e.g., poly[d(T)], poly[d(U)], and poly[d(C)]) than to others (e.g., poly[d(A)]; Newport et al., 1981; Pörschke and Rauh, 1983; Bulsink et al., 1985; Overman et al., 1988; Sang and Gray, 1989a; Lohman and Bujalowski, 1990; Ferrari and Lohman, 1994). The measured binding affinity of g5p for poly[r(U)] is greater than its affinity for poly[r(A)] at 5°C, but the binding preferences are calculated to be reversed at 37°C (Bulsink et al., 1985). From ^{1}H NMR studies by Alma et al. (1983b), a specific overlap of Phe-73 with the five-membered ring of bound $d(A)_{25–30}$ was proposed. Other NMR studies also showed the interaction of aromatic residues (Tyr-26 and Phe-73) in the DNA binding loop of the protein with bound nucleotides (King and Coleman, 1988; Folkers et al., 1993). The binding site has recently been modeled using the crystallographic structure (Guan et al., 1995), and NMR data have been used to model the site with a bound oligonucleotide tetramer (Edwards et al., 1997). These studies all suggest that there is some direct interaction between residues of the protein and the bases of bound nucleotides, which are perturbed from their single-stranded conformation. A question is whether such interactions confer binding preferences for certain bases.

The binding site of each g5p monomer accommodates four nucleotides in the most stable binding mode (Kansy et al., 1986). The g5p lends itself to a nearest-neighbor test of the influence of nucleic acid sequence on binding affinity, because of the protein's small binding site size and large protein dimer-dimer cooperativity factor. Homopolymers and simple sequences of three or five nucleotide repeats readily sample the possible positions of each nearest-neighbor within the binding site, while retaining the sensitivity of nearest-neighbor effects provided by simple sequence repeats. We chose to study ssDNA sequences because (a) an important function of g5p is to form a specific complex with the circular ssDNA phage genome, (b) the g5p is a model for other ssDNA-binding proteins, (c) differences in binding affinities for ssDNA cover a large range of experimental values, and (d) ssDNA is less subject to degradation than ssRNA. With this system, we have been able to show that differences in the binding affinity for DNA sequences of A and C bases are nearest-neighbor in their origin. This implies that specific protein-base interactions may be less important in determining sequence-dependent binding affinities than interactions between nucleotides in the free nucleic acid.

MATERIALS AND METHODS

Protein and oligomers

Wild-type Ff g5p was isolated from *E. coli* infected with fd virus as in previous work (Gray, 1989; Thompson et al., 1998). The purity was >96–99%, as determined by digital analysis of bands stained with Coomassie Blue on 18% sodium dodecyl sulfate-polyacrylamide gels. Protein concentrations were determined from absorption measurements using a molar extinction coefficient $\epsilon(276)$ of 7074 $M^{-1} \cdot cm^{-1}$ (Day, 1973). Six oligomers, each 48 nucleotides long, were purchased from Oligos Etc. (Wilsonville, OR). The sequences were dA_{48}, dC_{48}, $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$. Their concentrations were determined from absorption measurements and $\epsilon(260)$ values of 12,030, 7,440, 10,600, 9,090, 9,820, and 10,400 $M^{-1} \cdot cm^{-1}$, respectively, determined as in Gray et al. (1995).

UV absorption and CD measurements

UV absorption spectra were measured in Cary model 118 (Varian Associates, Houston, TX) and Olis-modified Cary model 14 (On-Line Instrument Co., Bogart, GA) spectrophotometers. CD spectra were measured using Jasco models J710 and J500A spectropolarimeters (Jasco, Easton, MD). Spectra were taken at 20 ± 0.5°C with samples in 1.0–2.0-cm path length cylindrical cells or rectangular cuvettes. Run scan speeds were 50 nm/min (J710) or 0.5 nm/s (J500A), the sensitivity was 50 mdeg/cm, response times were generally 1 s, and spectra with the J710 were the averages of four to six accumulations. Spectropolarimeters were calibrated as described by Gray et al. (1995). CD spectral data from the J710 were collected at 0.1-nm intervals and smoothed by the Savitzky and Golay (1964) method over 99 points with a third-order function. Data from the J500A were collected at 0.1-nm intervals and then subjected to two rounds of averaging by a seven-point quadratic function; spectra containing every 10th point were finally smoothed by a 13-point quadratic-cubic function (Savitzky and Golay, 1964). Baselines were subtracted before Savitzky-Golay smoothing. The smoothed CD data were plotted at 1-nm intervals as $\epsilon_L - \epsilon_R$ in units of $M^{-1} \cdot cm^{-1}$, per mol of nucleotide.

Titrations of oligomers with g5p

Titrations of different oligomers with gene 5 protein were performed as described by Kansy et al. (1986). The buffer was 2 mM Na^+ (phosphate), 5 mM Tris-HCl, and 0.1 M NaCl (pH 7.6), except that titrations of dA_{48} were at a lower NaCl concentration of 0.05 M. Final nucleotide concentrations were $4–7 \times 10^{-5}$ M in titrated g5p · oligomer complexes. The volume of added g5p was calculated by weighing before and after each addition.

Salt dissociations

Complexes formed at 1:4 protein monomer/nucleotide ratios in the above buffers, with g5p concentrations ranging from 3.2 to 20.1 μM, were dissociated by increasing the salt concentration with weighed aliquots of a 4 M NaCl solution. The density of the salt solution was taken into account in calculating salt concentrations and sample dilution. Corrections to the plotted dissociation data to account for protein dilution during salt titrations, relative to the protein concentration at the 50% dissociation value, were imperceptible. Light scattering due to protein precipitation during salt dissociations was minimal. The UV absorption at 320 nm for complexes (at a [protein monomer]/[nucleotide] molar ratio, [P]/[N], of 0.25 and at the highest protein and salt concentrations) ranged from 4.3 to 7.5% of the absorption at 260 nm of the respective complexes in the starting buffers.

The fraction of dissociated complex was monitored by CD spectroscopy. CD values at 270 or 280 nm for free oligomers in 2 mM Na^+ (phosphate), 5 mM Tris-HCl, with 1.0 M or 1.5 M NaCl were taken as reference values for completely dissociated complexes. The percent disso-

ciation was equal to 100 × {CD(free oligomer at high [NaCl]) − CD-(complex at given [NaCl])}/{CD(free oligomer at high [NaCl]) − CD-(saturated complex)}. Dissociation curves, plotted as percent dissociation (Y) as a function of [NaCl] (X), were fitted by the following nonlinear function, using a Marquardt-Levenberg algorithm as programmed in Sigmaplot (Jandel Scientific Software, San Rafael, CA):

$$Y = 100 \times \{1 - 1/[1 + (X/C)^B]\}. \quad (1)$$

C is the desired [NaCl] at the midpoint of the dissociation curve, and B is the slope at the midpoint. Standard errors of the fits were less than 5%. This fitting procedure was only used to determine the $[NaCl]_{50}$ concentration needed for 50% dissociation. The free protein concentration L_{50} at this $[NaCl]_{50}$ was related to an apparent binding affinity $K\omega_{app}$ as discussed in the next section.

THEORY

Binding to finite lattices

Salt dissociation curves, obtained by monitoring the perturbation of the nucleic acid CD, were highly cooperative for complexes formed with the 48-mer DNA oligomers (see Fig. 1). The finite lattice binding model of Epstein (1978) was used in two sets of calculations to compare with the measured dissociation data. In the first set, the dissociation was assumed to be all or none for a given lattice. In the second set, it was assumed that each lattice contained only one cluster of bound proteins, where the cluster could be of any size and at any position on the lattice, and that dissociation occurred from the ends of the cluster. Lattice lengths were taken to be either 24 or 48 nucleotides, because g5p dimers could bind to two linear oligomers or to one oligomer folded into a hairpin. The binding site per g5p dimer was taken to be four nucleotides on each of two strands, following the simultaneous dimer binding model described by Terwilliger (1996). For the relatively simple all-or-none and one-cluster models in which the number of cooperative dimer-dimer interactions is fixed relative to the number of bound dimers per lattice, the fraction of occupied sites, θ, could be calculated for different values of the product $K\omega L$. K is the intrinsic binding constant per dimer, ω is the cooperativity parameter for binding of adjacent g5p dimers, and L is the concentration of free, nonbound g5p dimers. Because the total protein concentration, L_{tot}, was fixed at a value sufficient to saturate the available sites, then $L = L_{tot}(1 - \theta)$, where $(1 - \theta)$ is the fraction of sites dissociated.

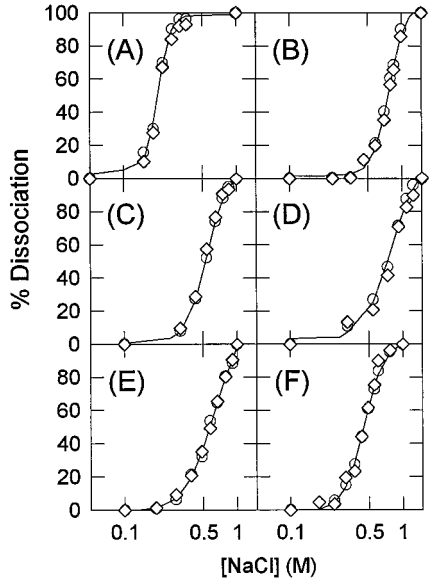

FIGURE 1 Representative salt dissociation curves of complexes formed between Ff g5p and different oligomers. Identical dissociation curves were obtained by monitoring CD values at 270 nm (○) and 280 nm (◇). The complexes and total g5p concentrations in these examples were (*A*) g5p · dA_{48} with 18.3 μM g5p; (*B*) g5p · dC_{48} with 20.1 μM g5p; (*C*) g5p · $d(AAC)_{16}$ with 18.6 μM g5p; (*D*) g5p · $d(ACC)_{16}$ with 17.2 μM g5p; (*E*) g5p · $d(AACC)_{12}$ with 20.0 μM g5p; and (*F*) g5p · $d(AAACC)_9A_3$ with 16.6 μM g5p. Data points were fitted by Eq. 1 as described in Materials and Methods to give the curves (——) from which 50% dissociation values were obtained. Data in this and the following figures were all taken at 20°C.

The θ was related to [NaCl] by using the equation $\log[K\omega] = c \log[NaCl] + b$, where c is the measured slope of the $\log[K\omega]/\log[NaCl]$ relationship and b is the intercept (Record et al., 1978). This equation allows one to relate the NaCl concentration, $[NaCl]_x$, for any $(K\omega)_x$ value to the $[NaCl]_{50}$ for the $(K\omega)_{50}$ value at 50% dissociation. Specifically,

$$[NaCl]_x = [NaCl]_{50}\{[(K\omega)_x/(K\omega)_{50}]^{1/c}\}$$
$$\equiv [NaCl]_{50}\{[((K\omega)_x L_{tot})/((K\omega)_{50} L_{tot})]^{1/c}\}.$$

Substituting $L_{tot} = L_x/(1 - \theta_x)$ and $L_{tot} = L_{50}/(1 - \theta_{50})$,

$$[NaCl]_x = [NaCl]_{50}\{[(K\omega L)_x(1 - \theta_{50})/((K\omega L)_{50}(1 - \theta_x))]^{1/c}\}. \quad (2)$$

That is, the calculated $(K\omega L)_x$ versus θ_x values [which include $(K\omega L)_{50}$ at θ_{50}] from the all-or-none or one-cluster models, together with the value of $[NaCl]_{50}$ from Eq. 1 and the value of the constant c, give the $[NaCl]_x$ concentration dependence of the fraction of sites dissociated $(1 - \theta_x)$ for a given complex. (The values of $K\omega$ from the two models differed by a constant factor at 50% dissociation, so the slopes c separately determined from $\log[K\omega_{app}]_{50}$ versus $\log[NaCl]_{50}$ plots were the same for all-or-none or one-cluster models.)

The calculated curves for different models are compared in Fig. 2 with the measured dissociation data for complexes with two of the oligomers. The data are all plotted relative to [NaCl] = 1 M at 50% dissociation. The extremes in breadth of the measured dissociation data are represented by the data for complexes with the repeating pentamer $d(AAACC)_9A_3$ and the repeating trimer $d(ACC)_{16}$. The all-or-none model closely approximates the dissociation data above 50% dissociation. Calculated curves for the all-or-none model with lattices of 24 and 48 nucleotides were very close on this scale; the curve shown is for the 48-mer lattice. The one cluster model fits the observed data for $d(ACC)_{16}$ better at the beginning of the dissociation. We interpret this to mean that the complexes essentially disso-

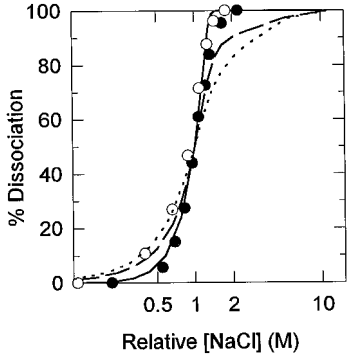

FIGURE 2 Comparison of measured salt dissociation data for g5p · $d(ACC)_{16}$ (○) and g5p · $d(AAACC)_9A_3$ (●) complexes with calculated curves for the all-or-none model with a 48-nucleotide lattice (——) and for the one cluster model with 48-nucleotide (— — —) and 24-nucleotide (- - -) lattices. The lattice in each case is assumed to have two strands for binding of a g5p dimer by the simultaneous binding model described by Terwilliger (1996). Plots are all at relative NaCl concentrations, scaled to [NaCl] = 1 M for 50% dissociation.

ciate in an all-or-none fashion except for any incompletely saturated lattices, or complexes formed between slightly misaligned DNA strands, which preferentially dissociate at the beginning of the salt titrations.

Hence, apparent values of the product, $K\omega_{app}$, per dimer were derived with the assumption that the dissociation was all or none by setting $K\omega_{app} = 1/(2L)$, where L was the free dimer concentration at 50% dissociation. These $K\omega_{app}$ values are those described by Terwilliger (1996) for the simultaneous binding model but are divided by 2 to correct for the twofold symmetry of the g5p (Thompson et al., 1998). This yields the $K\omega_{app}$ for binding of the g5p dimer in one orientation to two DNA strands (McGhee and von Hippel, 1974; Terwilliger, 1996; Thompson et al., 1998).

The binding constant for binding of m g5p dimers to a finite lattice is $K(K\omega)^{m-1}$, and the binding constant for m dimers in the midst of an infinite saturated polymer lattice is $(K\omega)^m$. Thus the derived values of $K\omega_{app}$ will be less than the values of $K\omega_{pol}$ for binding to polymers, because, for m adjacent dimers,

$$(K\omega_{app})^m = K(K\omega_{pol})^{m-1}.$$

Therefore,

$$(K\omega_{app})/(K\omega_{pol}) = \omega_{app}/\omega_{pol} = (\omega_{pol})^{-1/m}. \quad (3)$$

The binding of g5p dimers to oligomers longer than 16 nucleotides may result in hairpin formation (Cheng et al., 1996). Therefore, because g5p binds to the 48-mer oligonucleotides in the present experiments with a nucleotide/protein monomer ratio of 4, there could be as few as six adjacent g5p dimers bound per 48-mer. This means that values of $K\omega_{app}$ will be less than those for binding to long polymers, by at most a factor of $K\omega_{app}/K\omega_{pol} = (\omega_{pol})^{-1/6}$, assuming that K is the same for binding to oligomers and polymers. If the 48-mer oligomers bind without hairpin formation, the factor will be $(\omega_{pol})^{-1/12}$. The latter seems to be the case, as shown below in comparisons of data for oligomers and long polymers in Table 2. In any case, the interrelation of binding affinities to oligomers of equal length by the equations below is not at all affected, if the values of ω_{app} are the same for binding to all the sequences involved.

Relationship of binding affinities for different sequences

If the binding affinity of g5p for a dependent polynucleotide of mixed base sequence depends on the nucleotide composition of the sequence, the free energy of binding to the dependent sequence can be expressed in terms of the binding to independent homopolynucleotide sequences, i:

$$\Delta G^\circ(\text{dependent sequence}) = \sum_i f_i \Delta G^\circ(\text{independent sequence } i), \quad (4)$$

where f_i is the fraction of nucleotide i in the dependent sequence. This relationship describes the binding of random polynucleotides by the T4 phage g32p single-stranded DNA-binding protein (Newport et al., 1981; von Hippel et al., 1982). Such binding does not depend on the juxtaposition of neighboring bases, and it will be termed "zero-neighbor" binding. The independent sequences need not be uniform homopolynucleotides, but they must have base compositions different from that of the dependent sequence, and there will be appropriate changes in the weighting factors, f_i. In terms of binding affinities $K\omega_{app}$ for the g5p dimer,

$$\log[K\omega_{app}(\text{dependent sequence})] = \sum_i f_i \log[K\omega_{app}(\text{independent sequence } i)]. \quad (5)$$

If the cooperativity parameter, ω_{app}, is the same for binding to all of the sequences involved in Eq. 5, ω_{app} will cancel and the relationship will only depend on the intrinsic binding constants, K.

$$\log[K(\text{dependent sequence})] = \sum_i f_i \log[K(\text{independent sequence } i)]. \quad (6)$$

Zero-neighbor and nearest-neighbor models compared by singular value decomposition

Two models were considered to explain the relative binding affinities of g5p to the six oligomer sequences. In the zero-neighbor model, the binding is dependent on base composition, and there are six equations like Eq. 5 in two unknown parameters, $\log[K\omega(A)]$ and $\log[K\omega(C)]$, that are the logarithms of binding constants to individual A and C bases. (For simplicity, the subscript "app" on ω is dropped in Eqs. 7–9.) The set of six equations to be solved according

to the zero-neighbor model was

$$\log[K\omega(A_{48})] = \log[K\omega(A)],$$

$$\log[K\omega(C_{48})] = \log[K\omega(C)],$$

$$\log[K\omega((AAC)_{16})] = 0.67 \log[K\omega(A)]) + 0.33 \log[K\omega(C)],$$

$$\log[K\omega((ACC)_{16})] = 0.33 \log[K\omega(A)] + 0.67 \log[K\omega(C)],$$

$$\log[K\omega((AACC)_{12})] = 0.50 \log[K\omega(A)], + 0.50 \log[K\omega(C)],$$

and

$$\log[K\omega((AAACC)_9A_3)] = 0.625 \log[K\omega(A)], + 0.375 \log[K\omega(C)]. \quad (7)$$

In the nearest-neighbor model, parameters for binding to at least three independent sequence combinations are needed to define the binding to all dependent sequences with two types of bases. The simplest set of parameters consists of log[$K\omega$(AA)], log[$K\omega$(CC)], and log[$K\omega$(AC + CA)]. These are the logarithms of binding constants to AA, CC, and the average of AC plus CA nearest neighbors in the middle of a DNA sequence. It was assumed in choosing this set that end effects may be ignored and that the nearest neighbors may be counted as 48 for a 48-mer (i.e., as if the sequences were closed), although there are only 47 nearest neighbors (or fewer if the oligomer forms a hairpin) that are actually bound by g5p. Without this assumption, more than three independent sequences are needed to apply the nearest-neighbor model. Gray (1997a) has reviewed the numbers of independent sequences needed to describe a nearest-neighbor property for polymers and oligomers having two or more types of bases. The use of only three independent sequences in the present work was justified by the results below.

Thus the set of six equations to be solved according to the nearest-neighbor model was

$$\log[K\omega(A_{48})] = \log[K\omega(AA)],$$

$$\log[K\omega(C_{48})] = \log[K\omega(CC)],$$

$$\log[K\omega((AAC)_{16})] = 0.33 \log[K\omega(AA)] + 0.67 \log[K\omega(AC + CA)], \quad (8)$$

$$\log[K\omega((ACC)_{16})] = 0.33 \log[K\omega(CC)] + 0.67 \log[K\omega(AC + CA)],$$

$$\log[K\omega((AACC)_{12})] = 0.25 \log[K\omega(AA)], + 0.25 \log[K\omega(CC)] + 0.50 \log[K\omega(AC + CA)],$$

$$\log[K\omega((AAACC)_9A_3)] = 0.438 \log[K\omega(AA)] + 0.188 \log[K\omega(CC)] + 0.375 \log[K\omega(AC + CA)].$$

Solutions to the sets of linear Eqs. 7 and 8 were obtained from least-squares fits of the unknown parameters (two for the zero-neighbor model or three for the nearest-neighbor model) to the measured values. From log[$K\omega$] versus log[NaCl] data, measured values were extrapolated to 0.2 M NaCl. Such sets of overdetermined equations may be solved by singular value decomposition (SVD) (Press et al., 1992; Gray, 1997b). In matrix notation, the above equations are $\mathbf{log[K\omega(Dep)]_i} = \mathbf{N_{ij}} \times \mathbf{log[K\omega(Indep)]_j}$, where the $[K\omega(\text{Dep})]_i$, $i = 1$–6, are the measured values of binding constants for the six dependent oligomer sequences, the $[K\omega(\text{Indep})]_j$ are the values for the $j = 2$ or 3 independent sequences, and N_{ij} are the numbers of occurrences of the jth independent sequence in the ith oligomer. SVD of $\mathbf{N}_{ij}/\boldsymbol{\sigma}_i$ gives $\mathbf{USV^t}$. (σ_i is the standard deviation on $\log[K\omega(\text{Dep})]_i$ from the linear regression of $\log[K\omega(\text{Dep})]_i$ versus log[NaCl].) **S** is a diagonal $j \times j$ matrix of singular values, which were all nonzero for the chosen sets of independent sequences. The columns of the $6 \times j$ matrix **U** are orthonormal basis vectors that together with the significant values of **S** and the corresponding rows of the $j \times j$ matrix $\mathbf{V^t}$ (where **t** denotes a transpose) provide the best least-squares approximation to the original matrix $\mathbf{N_{ij}/\sigma_i}$. The values of $\log[K\omega(\text{Indep})]_j$ were obtained from $\mathbf{log[K\omega(Indep)]_j} = \mathbf{(VS^+U^t)} \times \mathbf{log[K\omega(Dep)]_i/\sigma_i}$, as described by others (Press et al., 1992; Compton and Johnson, 1986). Finally, substituting these values in Eqs. 7 and 8 gave the best-fit zero-neighbor and nearest-neighbor values for the six values of log[$K\omega$(Dep)], from which calculated $K\omega$(Dep) were obtained. These best-fit values could be compared with the original $K\omega$(Dep) values (see Table 1).

Error analysis

A statistical test of the fit of the models is given by the probability Q (Press et al., 1992) that the χ^2 value (where $\chi^2 = \Sigma_i|(\log[K\omega_{meas}]_i - \log[K\omega_{calc}]_i)/\sigma_i|^2$) from a fit to the data would be larger by chance. A larger Q indicates a better fit. Values of $Q > 0.1$ are adequate and may be acceptable above 0.001 (Press et al., 1992).

Errors tabulated for $K\omega$ were calculated from the errors on log[$K\omega$], which were either the measured standard deviations, σ, from linear regressions of $\log[K\omega(\text{Dep})]_i$ versus log[NaCl] or were the calculated standard deviations from the variance/covariance matrix of the SVD. As derived by Bevington and Robinson (1992), the errors, σ, in $K\omega$ and log[$K\omega$] are related as follows: $\sigma(K\omega) = \sigma(\ln[K\omega]) \div \partial(\ln[K\omega])/\partial(K\omega) = 2.303 \times (K\omega) \times \sigma(\log[K\omega])$.

Models compared by simple zero-neighbor and nearest-neighbor equations

The SVD of Eqs. 7 or 8 above is a means of deriving a simultaneous best fit to measured values of log[$K\omega$] for all six oligomers. However, the value of log[$K\omega$] for any one individual (dependent) sequence can be expressed in terms of the values for the remaining five (independent) sequences. Many combinations of independent sequences are possible, because only two or three of the remaining five are

needed for the zero-neighbor or nearest-neighbor approximations, respectively. We chose to use just dA_{48} and dC_{48} as the independent sequences for zero-neighbor calculations of $\log[K\omega]$ values of the remaining four sequences $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$. The combinations we chose for nearest-neighbor calculations for these four sequences were as follows:

$$\log[K\omega((AAC)_{16})] = \{3\log[K\omega((ACC)_{16})] + \log[K\omega(A_{48})] - \log[K\omega(C_{48})]\}/3,$$

$$\log[K\omega((ACC)_{16})] = \{3\log[K\omega((AAC)_{16})] + \log[K\omega(C_{48})] - \log[K\omega(A_{48})]\}/3,$$

$$\log[K\omega((AACC)_{12})] = \{3\log[K\omega((AAC)_{16})] + 3\log[K\omega((ACC)_{16})] + \log[K\omega(A_{48})] + \log[K\omega(C_{48})]\}/8,$$

$$\log[K\omega((AAACC)_9A_3)] = \{4\log[K\omega((AACC)_{12})] + 3\log[K\omega((AAC)_{16})] + 3\log[K\omega((ACC)_{16})] + 5\log[K\omega(A_{48})], + \log[K\omega(C_{48})]\}/16. \quad (9)$$

As for the SVD analysis, it was assumed that end effects could be ignored (i.e., that the sequences were closed). In Eqs. 9, each repeating trinucleotide sequence was chosen to be dependent on dA_{48} and dC_{48} and the other repeating trinucleotide. Equations for the repeating tetranucleotide and pentanucleotide sequences expressed the nearest neighbors in terms of all possible simpler sequences, weighted so that the AC plus CA neighbors in the independent sequences were equally represented.

Analogous equations can be set up in terms of any nearest-neighbor property. Equations analogous to Eq. 9 were used to calculate CD spectra of the free and complexed dependent oligomer sequences.

RESULTS

CD changes during formation and NaCl dissociation of complexes

Titrations of five of the DNA oligomers, dC_{48}, $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$, with wild-type Ff g5p were carried out in a buffer of 2 mM Na^+ (phosphate), 5 mM Tris-HCl, and 0.1 M NaCl, pH 7.6. The slightly elevated pH and 0.1 M NaCl were needed to eliminate the self-complexed form of dC_{48}. The CD spectrum of free dC_{48} (Fig. 3 *B*, *inset*) was that of single-stranded dC_{48}. Sequences with hemiprotonated $C^+ \cdot C$ base pairs have large positive and negative CD bands at 286 and 265 nm, respectively, that were absent under these conditions (Gray et al., 1988). The buffer for titration of dA_{48} contained only 0.05 M NaCl, because the binding affinity was weakest for this oligomer and stoichiometric binding was desired. All of the data were taken at 20°C.

The insets to Fig. 3 show representative spectra from titrations of each of the six DNA oligomer sequences with g5p. The addition of g5p caused the positive CD bands of the nucleic acids at wavelengths above 260 nm to decrease in magnitude or become negative. On this scale, the g5p has negligible CD at wavelengths above 250 nm, and the CD in this region reflects changes involving the DNA base chromophores. The DNA CD changes result from the change in nucleic acid conformation needed to span the binding sites in the interior of a left-handed superhelical structure (Gray, 1989; Olah et al., 1995; Thompson et al., 1998). Although the overall superhelical structure is left-handed, there could be residual right-handed base stacking, as inferred from the CD of complexes formed with poly[r(A)] (Sang and Gray, 1989b). Specifically, the changes in the DNA bands at long wavelengths (Fig. 3, *insets*) are reminiscent of dehydration effects on dsDNA (Johnson, 1996; Greve et al., 1987a,b). At shorter wavelengths, CD bands of g5p and DNA overlap. A tyrosyl CD band at 229 nm increases in magnitude because of the additions of g5p, but the relative increase is not as great as would be predicted from the known amount of added g5p, because of the perturbation of a single tyrosine (Tyr-34) at the interface of cooperatively bound proteins (Day, 1973; Kansy et al., 1986; Mark et al., 1995; Thompson et al., 1998). Making use of a Y34F mutant protein, we found that CD bands of poly[d(A)] and fd ssDNA below 250 nm are perturbed as if the bases were partially unstacked by heating to 70–80°C (Mark and Gray, 1997). Thus changes in the nucleic acid upon binding g5p are qualitatively similar to the combined effects of dehydration plus heating, with no obvious CD contributions from interactions between the DNA bases and protein chromophores at wavelengths above 210 nm (Mark and Gray, 1997).

The changes in absorbance at 260 nm upon forming complexes at [protein monomer]/[nucleotide] molar ratios, [P]/[N], of ~1/4 with dA_{48}, dC_{48}, $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$ were, respectively, 12.8, −3.8, 3.6, −4.2, 1.2, and 3.8% (with a maximum error range of ±1.9%). These values were determined as 100 × [OD(complex) − OD(free nucleic acid) − OD(free protein)]/[OD(free nucleic acid) + OD(free protein)]. The *hyper*chromicity of the d(AA)-containing oligomers and *hypo*chromicity or reduced hyperchromicity of d(CC)-containing and d(AC + CA)-containing oligomers were consistent with previous results that poly[d(A)] is 17.5–21.7% hyperchromic, whereas the pyrimidine polymers poly[d(C)] and poly[d(AC)] are both hypochromic (10.8% and 7.8%, respectively) upon complex formation with g5p (Sang and Gray, 1989a). By this measure, neighboring d(A) bases interacted less, whereas neighboring d(C) and alternating d(AC) bases interacted more in the bound state than in the free oligomers. These absorption data are in agreement with the CD data that the free d(C)-rich oligomers were initially single-stranded and less stacked than the free d(A)-rich sequences and that after binding there was residual interaction between neighboring bases.

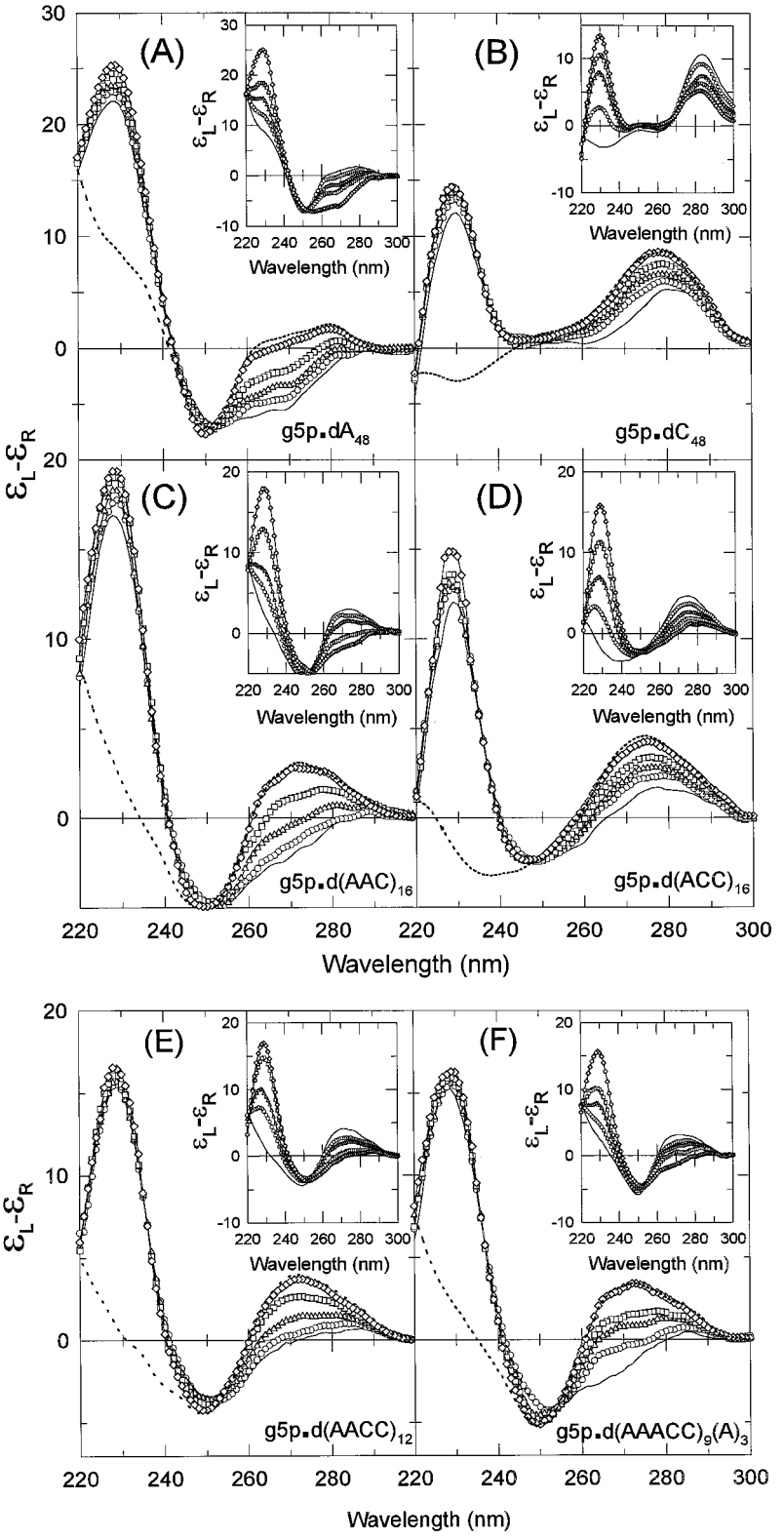

FIGURE 3 Main panels show representative CD spectra during example salt dissociations of the various g5p · oligomer complexes formed at [P]/[N] ratios of 1:4. Insets show representative CD spectra during titrations of the oligomers with g5p. The starting buffer was 2 mM Na^+ (phosphate), 5 mM Tris-HCl, 0.1 M NaCl, pH 7.6, for all of the oligomers except dA_{48}, for which the starting buffer contained 0.05 M instead of 0.1 M NaCl. (*A*) Main panel: Salt dissociation of g5p · dA_{48}; NaCl concentrations were 0.05 M (——), 0.10 M (○), 0.12 M (△), 0.15 M (□) and 0.22 M (◇). The CD spectrum of free dA_{48} at 1.0 M NaCl (- - - -) is shown for reference. The [g5p] was 7.0 μM in this sample. (*Inset*) CD titration of dA_{48} with g5p; [P]/[N] ratios were 0.00 (——), 0.05 (○), 0.10 (△), 0.15 (□), and 0.25 (◇). (*B*) Main panel: Salt dissociation of g5p · dC_{48}; NaCl concentrations were 0.10 M (——), 0.50 M (○), 0.67 M (△), 0.83 M (□), and 1.10 M (◇). The CD spectrum of free dC_{48} at 1.5 M NaCl (- - - -) is shown for reference. The [g5p] was 8.2 μM in this sample. (*Inset*) CD titration of dC_{48} with g5p; [P]/[N] ratios were 0.00 (——), 0.10 (○), 0.19 (△), 0.22 (□), and 0.26 (◇). (*C*) Main panel: Salt dissociation of g5p · $d(AAC)_{16}$; NaCl concentrations were 0.10 M (——), 0.24 M (○), 0.38 M (△), 0.48 M (□), and 0.64 M (◇). The CD spectrum of free $d(AAC)_{16}$ at 1.0 M NaCl (- - - -) is shown for reference. The [g5p] was 9.3 μM in this sample. (*Inset*) CD titration of $d(AAC)_{16}$ with g5p; [P]/[N] ratios were 0.00 (——), 0.05 (○), 0.10 (△), 0.19 (□), and 0.25 (◇). (*D*) Main panel: Salt dissociation of g5p · $d(ACC)_{16}$; NaCl concentrations were 0.10 M (——), 0.56 M (○), 0.67 M (△), 0.78 M (□), and 1.03 M (◇). The CD spectrum of free $d(ACC)_{16}$ at 1.5 M NaCl (– – –) is shown for reference. The [g5p] was 11.4 μM in this sample. (*Inset*) CD titration of $d(ACC)_{16}$ with g5p; [P]/[N] ratios were 0.00 (——), 0.07 (○), 0.13 (△), 0.19 (□), and 0.26 (◇). (*E*) Main panel: Salt dissociation of g5p · $d(AACC)_{12}$; NaCl concentrations were 0.10 M (——), 0.29 M (○), 0.48 M (△), 0.66 M (□), and 0.92 M (◇). The CD spectrum of free $d(AACC)_{12}$ at 1.0 M NaCl (- - - -) is shown for reference. The [g5p] was 12.3 μM in this sample. (*Inset*) CD titration of $d(AACC)_{12}$ with g5p; [P]/[N] ratios were 0.00 (——), 0.09 (○), 0.14 (△), 0.23 (□), and 0.27 (◇). (*F*) Main panel: Salt dissociation of g5p · $d(AAACC)_9A_3$; NaCl concentrations were 0.10 M (——), 0.23 M (○), 0.37 M (△), 0.44 M (□), and 0.89 M (◇). The CD spectrum of free $d(AAACC)_9A_3$ at 1.0 M NaCl (- - - -) is shown for reference. The [g5p] was 11.9 μM in this sample. (*Inset*) CD titration of $d(AAACC)_9A_3$ with g5p; [P]/[N] ratios were 0.00 (——), 0.05 (○), 0.10 (△), 0.15 (□), and 0.26 (◇). CD data in Figs. 3, 4, and 6 are plotted as ϵ_L-ϵ_R in units of $M^{-1} \cdot cm^{-1}$, per mole of nucleotide.

The main panels of Fig. 3 show representative CD spectra during salt dissociations of the six g5p · oligomer complexes. The complexes were all initially formed at a [P]/[N] ratio of 0.25. At increasing NaCl concentrations, the complexes dissociated, so the spectra above 260 nm had restored positive bands and became essentially like those of the respective free oligomers at comparably high salt concentrations (i.e., the dashed curves). The 229-nm band of the complexes also increased as the complexes dissociated, because of the relaxation of the perturbation of Tyr-34 (Fig.

3, *main panels*; compare the *solid curves* with the *curves with symbols*).

Changes in the nucleic acid CD at wavelengths above 260 nm were used to monitor the binding and salt dissociations.

Binding stoichiometry

CD titrations of oligomers were performed in the presence of 0.05 or 0.1 M NaCl, as described above. Plots of the changes in the long-wavelength CD bands of the oligomers at increasing [P]/[N] ratios are given in Fig. 4. Changes in the CD bands of the oligomers were essentially stoichiometric and proportional to the amount of protein added, until endpoints were reached at [P]/[N] ratios of 0.24–0.27. At the NaCl concentrations of 0.05–0.10 M used in these titrations of oligomers, only one endpoint was observed, indicative of the $n = 4$ binding mode (n is the number of nucleotides bound per g5p monomer). In previous work, CD titrations of poly[d(A)] and fd ssDNA in 2 mM sodium (phosphate), pH 7.0, revealed an additional binding mode at higher [P]/[N] ratios (Kansy et al., 1986; Mark et al., 1995), but this mode was absent in the present titrations of oligomers at a higher salt concentration (Fig. 4). Titration endpoints for the oligomers are summarized in Table 1.

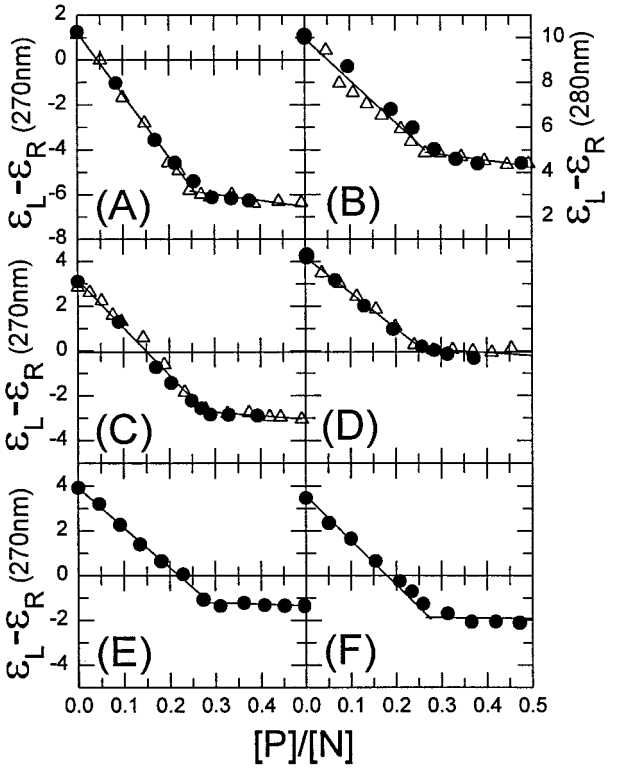

FIGURE 4 Titration plots showing the change in nucleic acid CD at long wavelengths upon addition of g5p. (*A*) dA_{48}. (*B*) dC_{48}. (*C*) $d(AAC)_{16}$. (*D*) $d(ACC)_{16}$. (*E*) $d(AACC)_{12}$. (*F*) $d(AAACC)_9A_3$. The CD at 270 nm is plotted for all of the oligomers except dC_{48}; for dC_{48} the value at 280 nm is plotted. Different symbols are data from separate titrations.

Dependence of $K\omega_{app}$ on [NaCl]

Complexes were formed at initial [P]/[N] ratios of 0.25 in the titration buffer described above, which contained 0.1 M NaCl except for dA_{48}, in which case it contained 0.05 M NaCl. Salt dissociation data are shown for one sample of each g5p · oligomer complex in Fig. 1. The total g5p concentrations in these samples were comparable, at 16–20 μM, and the NaCl concentrations needed for 50% dissociation varied from 0.25 to ~0.75 M. The g5p · dA_{48} complex dissociated at the lowest [NaCl] (Fig. 1 *A*), whereas complexes with dC_{48} and $d(ACC)_{16}$ dissociated at the highest [NaCl] (Fig. 1, *B* and *D*). Dissociation data were fitted with a function as described in Materials and Methods, from which NaCl concentrations at 50% dissociation values were obtained for complexes formed over a range of protein concentrations.

Fig. 5 shows that the logarithm of the binding constant, $\log[K\omega_{app}]$, was linearly related to the logarithm of the salt concentration, log[NaCl], for each complex. The data for g5p · dA_{48} and g5p · dC_{48} complexes (*open symbols*) are included in each panel for reference. Fig. 5, *A–D*, respectively, shows data (*closed symbols*) for complexes formed with $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$. The slopes of the $\log[K\omega_{app}]$ versus log[NaCl] plots are listed in the second column of Table 1. These slopes give the number of ions released per dimer, as explained by Terwilliger (1996). The slopes are the constants c used in Eq. 2. The values of $\log[K\omega_{app}]$/log[NaCl] for binding to these oligomers were similar and ranged from −3.0 to −3.4. They were comparable to the slope of −3.3 for binding to M13 phage ssDNA (Bulsink et al., 1985), but less than that of −4.65 to −4.7 for binding to poly[d(A)] (Bulsink et al., 1985; Terwilliger, 1996).

Of importance for the test of nearest-neighbor dependence is the fact that the experimental $K\omega_{app}$ values for binding to the six oligomers were significantly different and varied by over two orders of magnitude. Values extrapolated to the same [NaCl] of 0.2 M are shown in the third column of Table 1.

Nearest-neighbor dependence of $K\omega_{app}$

The $\log[K\omega_{app}]$ values for each of the four sequences that contained both A and C bases were calculated either using the average of the values for dA_{48} and dC_{48} (for zero-neighbor calculations) or using Eqs. 9 (i.e., nearest-neighbor equations) as described under Theory. Values of $\log[K\omega_{app}]$ for the dependent sequences were calculated from values of the independent sequences at the same NaCl concentrations, and the resulting points described straight lines with slopes that were the weighted averages of the slopes of the sequences whose $\log[K\omega_{app}]$ values were combined. The values calculated according to the zero-neighbor and nearest-neighbor models were very different, as seen by the

TABLE 1 Summary of CD titration endpoints, log[$K\omega_{app}$]/log[NaCl] dependence, and experimental and calculated $K\omega_{app}$ values for complexes of wild-type gene 5 protein and simple DNA sequences at 0.2 M NaCl, 20°C*

Sequence	Titration endpoint[#] [P]/[N]	Slope[§] $-\log[K\omega_{app}]/\log[NaCl]$	Experimental $K\omega_{app}$ ($\times 10^{-5}$ M^{-1})	Calculated $K\omega_{app}$ ($\times 10^{-5}$ M^{-1})[¶]	
				Zero-neighbor	Nearest-neighbor
dA_{48}	0.24 ± 0.02	3.23 ± 0.39	1.1 ± 0.07	1.6 ± 0.2	1.1 ± 0.2
dC_{48}	0.27 ± 0.02	3.44 ± 1.12	158 ± 23	739 ± 122	155 ± 45
$d(AAC)_{16}$	0.27 ± 0.01	3.00 ± 0.37	23.8 ± 2.0	12.5 ± 0.7	23.3 ± 2.6
$d(ACC)_{16}$	0.26 ± 0.01	3.29 ± 0.69	144 ± 13	96.6 ± 8.2	121 ± 11
$d(AACC)_{12}$	0.27 ± 0.01	3.01 ± 0.22	34.2 ± 1.1	34.8 ± 2.0	37.3 ± 2.2
$d(AAACC)_9A_3$	0.26 ± 0.02	3.23 ± 0.42	18.5 ± 1.0	16.2 ± 0.9	15.5 ± 0.9

*Experimental $K\omega_{app}$ data at other salt concentrations were extrapolated to 0.2 M NaCl.
[#]Titration endpoints and standard deviations were determined by fitting CD titration data at 270 nm with two linear regressions.
[§]Slope and standard deviations from linear regressions of the data in Fig. 5.
[¶]Zero-neighbor and nearest-neighbor values are from singular value decompositions and with standard deviations derived as described in the text.

separation of the dashed and dashed-dotted lines in Fig. 5. Only the values calculated according to the nearest-neighbor model were in agreement with the measured points.

The zero-neighbor and nearest-neighbor models were also tested by simultaneously fitting the data from all six sequences, using singular value decomposition to solve the sets of Eqs. 7 and 8 in the Theory section. SVD was carried out for values extrapolated to both 0.2 and 1.0 M NaCl. The results are tabulated in the last two columns of Table 1 for the values at 0.2 M NaCl. That is, the values for the six oligomer sequences in these columns are those whose logarithms are in perfect agreement according to the respective models. The calculated values were decidedly in favor of the nearest-neighbor model. The significance of the agreement for the nearest-neighbor model was provided by the statistical probability Q (see Theory) for fits of calculated to measured $\log[K\omega_{app}]$. Q was 0.36 with the nearest-neighbor model but was negligibly small ($<10^{-8}$) with the zero-neighbor model used to fit the logarithms of experimental $K\omega_{app}$ values shown in Table 1 at 0.2 M NaCl. The results were also significant for fits of data extrapolated to 1.0 M NaCl (not shown), with Q values of 0.81 and $<10^{-15}$ for fits with the nearest-neighbor and zero-neighbor models, respectively.

The nearest-neighbor agreement among the $\log[K\omega_{app}]$ values for all six oligomers supported the notion that g5p bound all of the oligomers in a similar fashion, reasonably described as all or none, and that nearest neighbors at the ends of the sequences could be neglected. The data were clearly not in agreement with the zero-neighbor model.

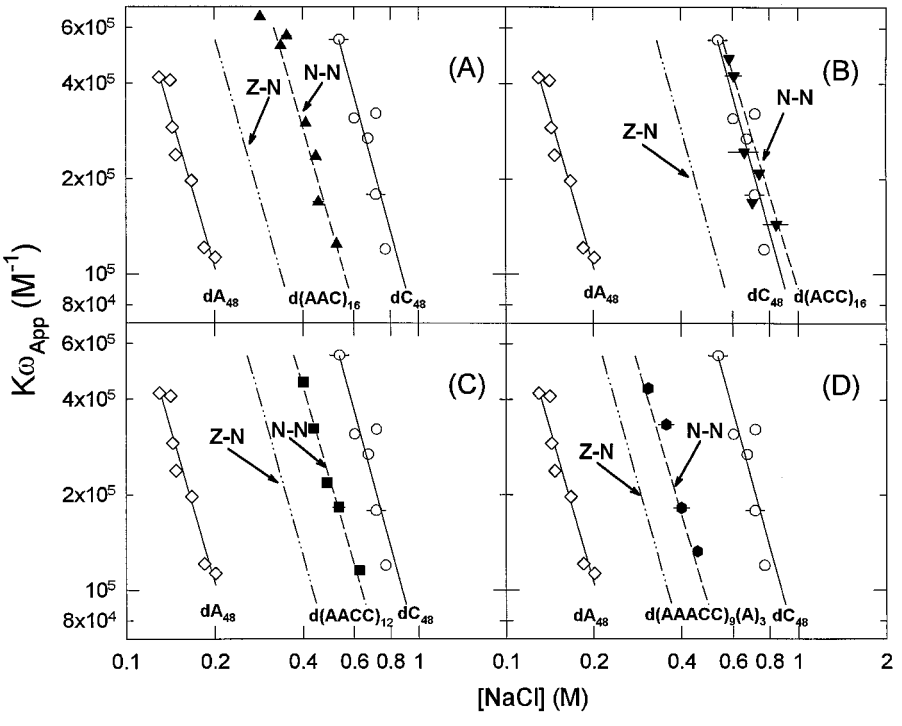

FIGURE 5 Test of the nearest-neighbor model for binding affinities to different sequences. The dependence of $\log[K\omega]$ on log[NaCl] for complexes with (*A*) $d(AAC)_{16}$ (▲), (*B*) $d(ACC)_{16}$ (▼), (*C*) $d(AACC)_{12}$ (■), and (*D*) $d(AAACC)_9A_3$ (●). Data for dA_{48} (◇) and dC_{48} (○) are shown in all four panels for reference. The binding affinities of the different oligomer sequences were consistent with nearest-neighbor (N-N) equations (– – –) but not zero-neighbor (Z-N) equations (– · · –).

Nearest-neighbor dependence of CD spectra

Because CD spectra are sensitive to the conformations of neighboring nucleotides, it was possible to use CD measurements to test whether the bound oligomers were in the same conformation. The CD spectra of free (*solid lines*) and complexed (*dashed lines*) oligomers are shown in Fig. 6. The spectra of the complexed oligomers include a large band at 229 nm from the g5p, but above 250 nm the spectral contribution of the protein is negligibly small. Hence, above 250 nm the spectra are dominated by the nucleic acid optical activity. The spectra of the free dA_{48} and dC_{48} oligomers shown in Fig. 6, *A* and *B*, differed and were differently perturbed by the g5p.

Nearest-neighbor equations analogous to Eq. 9 were used to calculate the spectra of the four oligomers $d(AAC)_{16}$, $d(ACC)_{16}$, $d(AACC)_{12}$, and $d(AAACC)_9A_3$, when free and when complexed with g5p. These spectra are shown in Fig. 6, *C–F*, along with the measured spectra. The CD spectrum of a nucleic acid is predominantly a nearest-neighbor property (Gray et al., 1992), and, consequently, spectra of the free oligomers (*solid lines*) in Fig. 6, *C–F*, were essentially identical to the spectra calculated from combinations of the spectra of other oligomers (*open circles*), as expected for sequences that were all in single-stranded conformations and devoid of any self-complexed structures. The measured spectra of the free oligomers did not agree with zero-neighbor calculations that were the weighted averages of the spectra of dA_{48} and dC_{48} (not shown).

If the average conformations of nearest neighbors in the bound sequences are the same for all of the sequences, with the CD contributions arising from interactions of the AA, CC, and AC plus CA nearest neighbors, the spectra of the complexes should obey the same nearest-neighbor relationships that hold for spectra of the free oligomers. This was the case, as seen by the comparisons of the measured (*dashed lines*) and calculated (*open triangles*) spectra in Fig. 6, *C–F*. The measured spectra for the complexes were not in agreement with simple zero-neighbor averages of the spectra of bound dA_{48} and dC_{48} (not shown). CD spectra are very sensitive to differences in average nucleotide geometry, yet the spectra of the bound oligomers were related by nearest-neighbor equations as precisely as were the spectra of the free oligomers. Therefore, CD spectra of the bound oligomer sequences implied that each type of nearest neighbor, regardless of the sequence, had a similar average structure within the binding site.

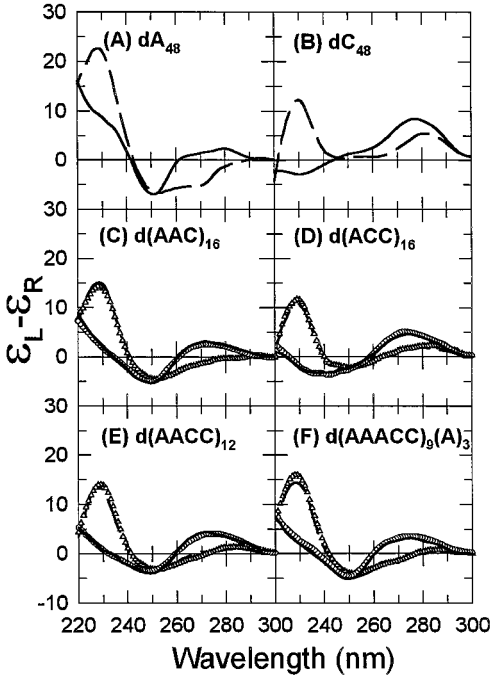

FIGURE 6 Test of the nearest-neighbor model for CD spectra of different sequences, free and complexed with g5p. (*A*) dA_{48}. (*B*) dC_{48}. (*C*) $d(AAC)_{16}$. (*D*) $d(ACC)_{16}$. (*E*) $d(AACC)_{12}$. (*F*) $d(AAACC)_9A_3$. In all panels, experimental spectra of free oligomers (——) and bound oligomers (– – –) are shown. In *C–F*, spectra calculated using nearest-neighbor equations for free (○) and bound (△) oligomers are shown. Standard deviations from measurements of five to eight samples of each of the complexes gave error bars (not shown) that were approximately the width of the symbols.

Estimates of $\omega_{app}/\omega_{pol}$

As was described in the Theory section, binding to oligomers results in a binding affinity $K\omega_{app}$, with a cooperativity factor that is less than the ω_{pol} for binding to polymers. This does not affect the nearest-neighbor relationships, as long as the value of ω_{app} is the same for all of the oligomer sequences and can be omitted from Eq. 5. However, the values of the ratio $\omega_{app}/\omega_{pol}$ can be estimated from our present data and from data previously presented by Sang and Gray (1989a) for binding to poly[d(A)], poly[d(C)], and poly[d(AC)]. The results are shown in Table 2. As expected, the $\omega_{app}/\omega_{pol}$ ratio is less than 1.0, indicating that the binding affinity for oligomers that are 48 nucleotides in length is less than that for binding to polymers. The ratio is calculated at 0.2 M NaCl for poly[d(A)] and at 1.0 M for the other two polymers to minimize errors from extrapolations of $\log[K\omega]/\log[\mathrm{NaCl}]$ plots. (Others have concluded that the value of ω is essentially independent of salt concentration (Alma et al., 1983a; Pörschke and Rauh, 1983; Terwilliger, 1996), so the ratio should be insensitive to salt concentration.) The ratio $\omega_{app}/\omega_{pol}$ equals $(\omega_{pol})^{-1/m}$, where m is the number of cooperatively bound monomers (see Theory). The ratio is more sensitive to m than to the value of ω_{pol}. The last columns of Table 2 show the values calculated for $\omega_{pol} = 500$ and 5000, covering the range of values reported by Bulsink et al. (1985) and Terwilliger (1996), and for 6 or 12 monomers bound per 48-mer oligomer. The conclusion is that the values of $K\omega_{app}$ we have measured are most consistent with the binding of two 48-mer oligomers to 12 g5p dimer binding sites, with each 48-mer oligomer saturated by 12 contiguous g5p monomers. This is in agreement with the values of saturation endpoints in Fig. 2 and Table 1, which are close to four nucleotides per g5p monomer.

TABLE 2 Measured $K\omega$ values for three g5p · DNA polymer complexes and values derived from complexes with oligomers

Sequence	Experimental for polymer* $K\omega_{pol}$ ($\times 10^{-3}$ M^{-1})	Nearest-neighbor calc. for oligomer# $K\omega_{app}$ ($\times 10^{-3}$ M^{-1})	Ratio $\omega_{app}/\omega_{pol}$	Calculated ratio for six dimers per oligomer $=(\omega_{pol})^{-1/6}$	Calculated ratio for 12 dimers per oligomer $=(\omega_{pol})^{-1/12}$
dA_n or dA_{48}	161 ± 10	110 ± 16	0.68 ± 0.10		
dC_n or dC_{48}	111 ± 7.5	63.6 ± 18.7	0.57 ± 0.17		
$d(AC)_n$ or $d(AC)_{24}$	136 ± 6.4	108 ± 20	0.80 ± 0.15		
Mean			0.68 ± 0.08§	0.24–0.35¶	0.49–0.60¶

*Measured values at 20°C are from published data at 0.2 M NaCl for poly[d(A)] and at 1.0 M NaCl for the other two polymers (Sang and Gray, 1989a). At these salt concentrations errors due to extrapolations are minimized.

#Nearest-neighbor calculated values are from singular value decompositions of data with six oligomers, derived as described in the text. The calculated values are those of the independent sequences defined by Eq. 8, extrapolated to the desired salt concentrations.

§Weighted standard deviation of the mean.

¶The values are calculated for $\omega_{pol} = 5000$ and 500, respectively.

DISCUSSION

The above results show that differences in the binding affinity of the g5p for different ssDNA sequences that have A and C bases are related to the nearest-neighbor arrangement of bases in the ssDNA sequence. Schematically, for a single strand of *n* nucleotides complexed with n/4 g5p monomers,

$$\ldots.C_{xx}A_xA_{xx}C_{xxx}C_{xx}A_xA_{xx}C.\ldots \times (n/4)(g5p) \underset{K\omega}{\longleftrightarrow} (g5p)_{n/4} \cdot (\ldots.C\text{-}A\text{-}A\text{-}C\text{-}C\text{-}A\text{-}A\text{-}C.\ldots), \quad (10)$$

where decreasing nearest-neighbor interactions in the free ssDNA are indicated by "x", "xx", and "xxx." When bound within the g5p binding site, the various nearest neighbors continue to interact, as shown by the nearest-neighbor dependence of the CD spectra of the bound sequences and hypochromicity of CC and AC + CA neighbors (present results and Sang and Gray, 1989a). The simplest model for explaining the present data is that higher binding affinities (under our solution conditions) are correlated with the degree of unstacking of the bases in the free DNA sequences, with each type of nearest-neighbor having approximately the same average base-base and base-protein interaction energies, regardless of the sequence, once the sequence is within the binding site. It should be noted that previous NMR studies showed that the H2 protons of $d(A)_n$ oligomers ($n = 25$–30 or 40–60) shifted to a common resonance upon binding that was consistent with a partially unstacked d(A) structure in which the bases were in rapid exchange between binding site positions (Alma et al., 1983b; King and Coleman, 1988). Such an averaged structure is schematically depicted by the similar distances between the bases in the right-hand side of Eq. 10. Fig. 7 illustrates this simplest model of the nearest-neighbor dependence of binding affinity.

This general notion is supported by other studies. Poly[d(A)] is more extensively stacked than is poly[d(C)] (Antao and Gray, 1993; other literature has been reviewed by Ferrari and Lohman, 1994). Enthalpic stacking energies have been calculated to be more favorable for d(ApA) than for d(CpC) and to be intermediate for d(CpA) in single strands of B-DNA (Gupta and Sasisekharan, 1978). Experimental enthalpies are greater for stacking of purine bases than for stacking of pyrimidine nucleosides in water (Saenger, 1984). Bulsink et al. (1985) previously interpreted the near-zero (±2 kcal/mol) enthalpy change for binding of g5p to poly[d(A)], compared with a value of $\Delta H^\circ_{obs} = -17.2$ kcal/mol protein monomer for binding to poly[d(U)], as the compensation of a favorable base-protein enthalpy interaction by the unstacking of bases in the case of poly[d(A)]. Ferrari and Lohman (1994) concluded that the binding of $dA(pA)_n$ by *Escherichia coli* SSB is accompanied by significant base unstacking, which results in a nonlinear van't Hoff plot not found for the binding of oligomers $dT(pT)_n$ and $dC(pC)_n$ that do not stack. The ΔH°_{obs} for binding of oligodeoxypyrimidines by *E. coli* SSB is more negative (favorable) than for binding of oligodeoxyadenylates, probably partly because of stacking of aromatic residues with bases that are already unstacked in the case of the oligodeoxypyrimidines (Ferrari and Lohman, 1994; Kozlov and Lohman, 1998). Our data for the Ff g5p are not only in agreement with these conclusions based on homopurine and homopyrimidine sequences; they provide new information that differences in stacking enthalpies of the nearest neighbors might dominate the sequence dependence of the binding of g5p to more complex sequences.

Sequence-dependent entropic effects could also play a role. For binding of g5p to sequences such as poly[d(A)] at a physiological salt concentration, where $\Delta H^\circ_{obs} \approx \pm 2$ kcal/mol, the binding affinity is then dominated by a favorable (positive) ΔS°_{obs} (Bulsink et al., 1985). The tabulated data presented by Bulsink et al. (1985) suggest that ΔS°_{obs} is unfavorable (i.e., negative) for binding of unstacked poly[d(U)] at the same salt concentration, if one relies on data extrapolated from a high salt concentration with the assumption that ΔH°_{obs} is not salt dependent.

The binding enthalpies summarized by Bulsink et al. (1985) were for the binding of g5p to poly[d(U)] and

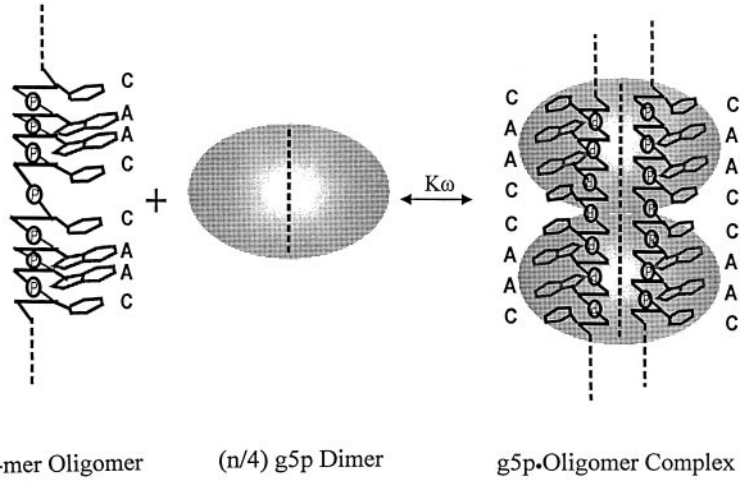

FIGURE 7 A simple model. Our results show that differences in binding affinity of the g5p to ssDNA sequences having A and C bases are related to the nearest-neighbor arrangement of bases in the ssDNA sequence. The origin of this effect could be the different nearest-neighbor interactions in the free ssDNA, indicated in this illustration by base-base distances that are smallest for A-A and largest for C-C. Thus unstacked C-C neighbors are more easily perturbed and bind more tightly than stacked A-A neighbors. A-C plus C-A nearest neighbors on average are also more easily perturbed than A-A neighbors. When bound within the g5p binding site, the various nearest neighbors continue to interact, as shown by the CD spectra of the bound sequences and the hypochromicity or reduced hyperchromicity values for C-C and A-C plus C-A neighbors. Once bound, the average interaction energies of the various types of nearest neighbor could be similar. (Only two adjacent dimers in the midst of a long sequence are depicted in this illustration.)

poly[d(A)] at different NaCl concentrations of 0.7 M and 0.175–0.21 M, respectively. Different salt concentrations can affect the values of ΔH°_{obs}. In the case of *E. coli* SSB, the binding to poly[U], $dA(pA)_{69}$, and $dT(pT)_{69}$ has a significant dependence on [NaCl], possibly caused by preferentially bound Cl^-, which results in a ΔH°_{obs} that is less negative with increasing [NaCl] (Lohman et al., 1996; Kozlov and Lohman, 1998). Detailed thermodynamic data are not available to similarly characterize the binding of different sequences by the Ff g5p. In the present experiments, the sequence-dependent binding affinities maintain the same dependence on nearest-neighbor compositions over a wide range of [NaCl] from 0.2 to 1 M (Fig. 5, Table 1, and text). Therefore, if there is an effect of [NaCl] on ΔH°_{obs} (the stacking of bases relative to interactions in the binding site), the effect is similar for all six oligomer sequences (assuming that changes in ΔS°_{obs} with NaCl concentration are due primarily to counterion release (Mascotti and Lohman, 1997; Lohman et al., 1996) and are the same for all of the oligomer sequences).

More complicated models would be consistent with a nearest-neighbor dependence of binding affinities. For example, each type of nearest neighbor (A-A, C-C, and the average of A-C plus C-A) could have a different average conformation and energy when bound, as long as the average remains essentially invariant for that nearest neighbor in different bound sequences. It should be noted that the conformation of a given nearest neighbor (in sequences containing A and C) appears to be approximately independent of its position within the binding site, given the nearest-neighbor agreement of the data for $d(AACC)_{12}$ with the data for the other sequences. This sequence should be phased in the g5p binding site when bound in the $n = 4$ binding mode. Because this sequence includes all four nearest neighbors, if any one were to be unusually perturbed or to contribute in a dramatically different manner to the CD or binding affinity, the data for this sequence would not be consistent with the data for the other sequences that are not phased in the binding site.

Whatever the details of the interactions, our data support the notion that the nearest-neighbor properties of the sequences themselves, at least for ssDNA sequences containing A and C, are major determinants of the binding preferences of g5p. Only three combinations of four types of nearest neighbors have been tested in this work. A total of 16 different nearest neighbors (neglecting ends) and 13 independent sequence combinations would have to be tested (Gray, 1997a) to determine whether g5p has the ability to discriminate among ssDNA sequences using determinants other than nearest-neighbor interactions. Moreover, the binding of g5p to special RNA sequences, important for the translational regulatory ability of the g5p (Zaman et al., 1991), could be more complex. Because neither Ff g5p nor T4 g32p represses translation of RNAs that are repressed by the other protein, it has been suggested that binding involves some specific sequence recognition (Fulford and Model, 1984). In fact, the binding to DNA, to inhibit replication, and to RNA, to inhibit translation, has been shown to involve different amino acid residues (Stassen et al., 1992). Other ions or solution conditions may affect how the protein binds to specific sequences. Therefore, the sequence discrimination of g5p may not always be dominated by the nearest-neighbor structural properties of the nucleic acid sequence.

For other proteins that bind single-stranded nucleic acids, the relative influence of stacking might be less than in the case of Ff g5p. The sequence dependence of the binding of the phage T4 g32p has been described as a zero-neighbor

(i.e., base composition-dependent) property (Newport et al., 1981), although this conclusion was drawn from binding data for randomly copolymerized polyribonucleotides or natural DNA, which were not designed to provide a clear distinction between the zero-neighbor and nearest-neighbor models. Natural RNA loading sites for the *E. coli* rho protein are unstructured and rich in cytosine residues, but base-paired RNA secondary structures may be particularly important in influencing the in vivo activity of the rho protein (Walstrom et al., 1998). The *E. coli* helicase DnaB protein has a higher affinity for poly[d(A)] than for poly[d(T)], suggesting an important contribution of the base to the binding affinity (Jezewska et al., 1996). In the case of the DNA-binding domain of the human replication protein A, for which a structure bound to dC_8 has been solved, there is evidence that there may be a plasticity in the binding and that alternative residues may be involved in binding to different ssDNA sequences (Bochkarev et al., 1997). The extent to which the stacked structure of the free nucleic acid influences the binding affinities of this and other ssDNA-binding proteins will depend on whether subsets of alternative interactions for a wide range of bound sequences average out to be of similar energies.

This work was performed by T.-C. M. in partial fulfillment of the requirements for the Ph.D. degree in the Dept. of Molecular and Cell Biology, The University of Texas at Dallas.

This work was supported by National Science Foundation grant MCB-9405683 and Robert A. Welch Foundation grant AT-503.

REFERENCES

Alberts, B., L. Frey, and H. Delius. 1972. Isolation and characterization of gene 5 protein of filamentous bacterial viruses. *J. Mol. Biol.* 68: 139–152.

Allison, T. J., T. C. Wood, D. M. Briercheck, F. Rastinejad, J. P. Richardson, and G. S. Rule. 1998. Crystal structure of the RNA-binding domain from transcription termination factor rho. *Nature Struct. Biol.* 5:352–356.

Alma, N. C. M., B. J. M. Harmsen, E. A. M. DeJong, J. v. d. Ven, and C. W. Hilbers. 1983a. Fluorescence studies of the complex formation between the gene 5 protein of bacteriophage M13 and polynucleotides. *J. Mol. Biol.* 163:47–62.

Alma, N. C. M., B. J. M. Harmsen, J. H. van Boom, G. van der Marel, and C. W. Hilbers. 1983b. A 500-MHz proton nuclear magnetic resonance study of the structure and structural alterations of gene-5 protein-oligo(deoxyadenylic acid) complexes. *Biochemistry.* 22:2104–2115.

Antao, V. P., and D. M. Gray. 1993. CD spectral comparisons of the acid-induced structures of poly[d(A)], poly[r(A)], poly[d(C)], and poly[r(C)]. *J. Biomol. Struct. Dyn.* 10:819–839.

Bauer, M., and G. P. Smith. 1988. Filamentous phage morphogenetic signal sequence and orientation of DNA in the virion and gene-V protein complex. *Virology.* 167:166–175.

Bevington, P. R., and D. K. Robinson. 1992. Data Reduction and Error Analysis for the Physical Sciences, 2nd Ed. McGraw-Hill, New York. 48.

Bochkarev, A., R. A. Pfuetzner, A. M. Edwards, and L. Frappier. 1997. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature.* 385:176–181.

Briercheck, D. M., T. C. Wood, T. J. Allison, J. P. Richardson, and G. S. Rule. 1998. The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA-protein interactions. *Nature Struct. Biol.* 5:393–399.

Bulsink, H., B. J. M. Harmsen, and C. W. Hilbers. 1985. Specificity of the binding of bacteriophage M13 encoded gene-5 protein to DNA and RNA studied by means of fluorescence titrations. *J. Biomol. Struct. Dyn.* 3:227–247.

Cheng, X., A. C. Harms, P. N. Goudreau, T. C. Terwilliger, and R. D. Smith. 1996. Direct measurement of oligonucleotide binding stoichiometry of gene V protein by mass spectrometry. *Proc. Natl. Acad. Sci. USA.* 93:7022–7027.

Compton, L. A., and W. C. Johnson, Jr. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 155:155–167.

Day, L. A. 1973. Circular dichroism and ultraviolet absorption of a deoxyribonucleic acid binding protein of filamentous bacteriophage. *Biochemistry.* 12:5329–5339.

Edwards, C. A., S. V. S. Mariappan, C.-S. Tung, and T. C. Terwilliger. 1997. Gene V protein-DNA interactions. *Biophys. J.* 72:A344.

Epstein, I. R. 1978. Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. *Biophys. Chem.* 8:327–339.

Ferrari, M. E., and T. M. Lohman. 1994. Apparent heat capacity change accompanying a nonspecific protein-DNA interaction. *Escherichia coli* SSB tetramer binding to oligodeoxyadenylates. *Biochemistry.* 33: 12896–12910.

Folkers, P. J. M., J. P. M. van Duynhoven, H. T. M. van Lieshout, B. J. M. Harmsen, J. H. van Boom, G. I. Tesser, R. N. H. Konings, and C. W. Hilbers. 1993. Exploring the DNA binding domain of gene V protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with ^{1}H-NMR. *Biochemistry.* 32:9407–9416.

Folmer, R. H. A., M. Nilges, P. J. M. Folkers, R. N. H. Konings, and C. W. Hilbers. 1994. A model of the complex between single-stranded DNA and the single-stranded DNA binding protein encoded by gene V of filamentous bacteriophage M13. *J. Mol. Biol.* 240:341–357.

Folmer, R. H. A., M. Nilges, R. N. H. Konings, and C. W. Hilbers. 1995. Solution structure of the single-stranded DNA binding protein of the filamentous *Pseudomonas* phage Pf3: similarity to other proteins binding to single-stranded nucleic acids. *EMBO J.* 14:4132–4142.

Folmer, R. H. A., M. Nilges, C. H. M. Papavoine, B. J. M. Harmsen, R. N. H. Konings, and C. W. Hilbers. 1997. Refined structure, DNA binding studies, and dynamics of the bacteriophage Pf3 encoded single-stranded DNA binding protein. *Biochemistry.* 36:9120–9135.

Fulford, W., and P. Model. 1984. Specificity of translational regulation by two DNA-binding proteins. *J. Mol. Biol.* 173:211-226.

Gray, C. W. 1989. Three-dimensional structure of complexes of single-stranded DNA-binding proteins with DNA. IKe and fd gene 5 proteins form left-handed helices with single-stranded DNA. *J. Mol. Biol.* 208: 57–64.

Gray, D. M., C. W. Gray, and R. D. Carlson. 1982. Neutron scattering data on reconstituted complexes of fd deoxyribonucleic acid and gene 5 protein show that the deoxyribonucleic acid is near the center. *Biochemistry.* 21:2702–2713.

Gray, D. M. 1997a. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors. *Biopolymers.* 42:783–793.

Gray, D. M. 1997b. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA · RNA hybrids and DNA duplexes. *Biopolymers.* 42:795–810.

Gray, D. M., S.-H. Hung, and K. H. Johnson. 1995. Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.* 246:19–34.

Gray, D. M., R. L. Ratliff, V. P. Antao, and C. W. Gray. 1988. CD spectroscopy of acid-induced structures of polydeoxyribonucleotides: importance of $C \cdot C^+$ base pairs. *In* Structure and Expression, Vol. 2, DNA and Its Drug Complexes. R. H. Sarma and M. H. Sarma, editors. Adenine Press, Schenectady, NY. 147–166.

Gray, D. M., R. L. Ratliff, and M. R. Vaughan. 1992. Circular dichroism spectroscopy of DNA. *Methods Enzymol.* 211:389–406.

Greve, J., M. F. Maestre, H. Moise, and J. Hosoda. 1978a. Circular dichroism study of the interaction between T4 gene 32 protein and polynucleotides. *Biochemistry.* 17:887–893.

Greve, J., M. F. Maestre, H. Moise, and J. Hosoda. 1978b. Circular dichroism studies of the interaction of a limited hydrolysate of T4 gene 32 protein with T4 DNA and poly[d(A-T)] · poly[d(A-T)]. *Biochemistry.* 17:893–898.

Guan, Y., H. Zhang, R. N. H. Konings, C. W. Hilbers, T. C. Terwilliger, and A. H.-J. Wang. 1994. Crystal structures of Y41H and Y41F mutants of gene V protein from Ff phage suggest possible protein-protein interactions in the GVP-ssDNA complex. *Biochemistry.* 33:7768–7778.

Guan, Y., H. Zhang, and A. H.-J. Wang. 1995. Electrostatic potential distribution of the gene V protein from Ff phage facilitates cooperative DNA binding: a model of the GVP-ssDNA complex. *Protein Sci.* 4:187–197.

Gupta, G., and V. Sasisekharan. 1978. Theoretical calculations of base-base interactions in nucleic acids: II. Stacking interactions in polynucleotides. *Nucleic Acids Res.* 5:1655–1673.

Heyer, W.-D., and R. D. Kolodner. 1989. Purification and characterization of a protein from *Saccharomyces cerevisiae* that binds tightly to single-stranded DNA and stimulates a cognate strand exchange protein. *Biochemistry.* 28:2856–2862.

Heyer, W.-D., M. R. S. Rao, L. F. Erdile, T. J. Kelly, and R. D. Kolodner. 1990. An essential *Saccharomyces cerevisiae* single-stranded DNA binding protein is homologous to the large subunit of human RP-A. *EMBO J.* 9:2321–2329.

Ikoku, A. S., and J. E. Hearst. 1981. Identification of a structural hairpin in the filamentous chimeric phage M13Gori1. *J. Mol. Biol.* 151:245–259.

Jezewska, M. J., U.-S. Kim, and W. Bujalowski. 1996. Binding of *Escherichia coli* primary replicative helicase DnaB protein to single-stranded DNA. Long-range allosteric conformational changes within the protein hexamer. *Biochemistry.* 35:2129–2145.

Johnson, W. C., Jr. 1996. Determination of the conformation of nucleic acids by electronic CD. *In* Circular Dichroism and the Conformational Analysis of Molecules. G. D. Fasman, editor. Plenum Press, New York. 433–468.

Kansy, J. W., B. A. Clack, and D. M. Gray. 1986. The binding of fd gene 5 protein to polydeoxynucleotides: evidence from CD measurements for two binding modes. *J. Biomol. Struct. Dyn.* 3:1079–1110.

King, G. C., and J. E. Coleman. 1988. The Ff gene 5 protein-d$(pA)_{40-60}$ complex: ^{1}H NMR supports a localized base-binding model. *Biochemistry.* 27:6947–6953.

Kozlov, A. G., and T. M. Lohman. 1998. Calorimetric studies of *E. coli* SSB protein-single-stranded DNA interactions. Effects of monovalent salts on binding enthalpy. *J. Mol. Biol.* 278:999–1014.

Lohman, T. M., and W. Bujalowski. 1990. *Escherichia coli* single-strand binding protein: multiple single-stranded DNA binding modes and cooperativities. *In* The Biology of Nonspecific DNA-Protein Interactions. A. Revzin, editor. CRC Press, Boca Raton, FL. 131–170.

Lohman, T. M., L. B. Overman, M. E. Ferrari, and A. G. Kozlov. 1996. A highly salt-dependent enthalpy change for *Escherichia coli* SSB protein-nucleic acid binding due to ion-protein interactions. *Biochemistry.* 35: 5272–5279.

Mark, B. L., and D. M. Gray. 1997. Tyrosine mutant helps define overlapping CD bands from fd gene 5 protein · nucleic acid complexes. *Biopolymers.* 42:337–348.

Mark, B. L., T. C. Terwilliger, M. R. Vaughan, and D. M. Gray. 1995. Circular dichroism spectroscopy of three tyrosine-to-phenylalanine substitutions of fd gene 5 protein. *Biochemistry.* 34:12854–12865.

Mascotti, D. P., and T. M. Lohman. 1997. Thermodynamics of oligoarginines binding to RNA and DNA. *Biochemistry.* 36:7272–7279.

McGhee, J. D., and P. H. von Hippel. 1974. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86: 469–489.

Michel, B., and N. D. Zinder. 1989. *In vitro* binding of the bacteriophage f1 gene V protein to the gene II RNA-operator and its DNA analog. *Nucleic Acids Res.* 17:7333–7344.

Newport, J. W., N. Lonberg, S. C. Kowalczykowski, and P. H. von Hippel. 1981. Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. II. Specificity of binding to DNA and RNA. *J. Mol. Biol.* 145:105–121.

Olah, G. A., D. M. Gray, C. W. Gray, D. L. Kergil, T. R. Sosnick, B. L. Mark, M. R. Vaughan, and J. Trewhella. 1995. Structures of fd gene 5 protein · nucleic acid complexes: a combined solution scattering and electron microscopy study. *J. Mol. Biol.* 249:576–594.

Overman, L. B., W. Bujalowski, and T. M. Lohman. 1988. Equilibrium binding of *Escherichia coli* single-strand binding protein to single-stranded nucleic acids in the $(SSB)_{65}$ binding mode. Cation and anion effects and polynucleotide specificity. *Biochemistry.* 27:456–471.

Pörschke, D., and H. Rauh. 1983. Cooperative, excluded-site binding and its dynamics for the interaction of gene 5 protein with polynucleotides. *Biochemistry.* 22:4737–4745.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C, 2nd Ed. Cambridge University Press, New York. 59–70.

Raghunathan, S., C. S. Ricard, T. M. Lohman, and G. Waksman. 1997. Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction of the selenomethionyl protein at 2.9-Å resolution. *Proc. Natl. Acad. Sci. USA.* 94:6652–6657.

Record, M. T., Jr., C. F. Anderson, and T. M. Lohman. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* 11:103–178.

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York. 135.

Sang, B.-C., and D. M. Gray. 1989a. Specificity of the binding of fd gene 5 protein to polydeoxyribonucleotides. *J. Biomol. Struct. Dyn.* 7:693–706.

Sang, B.-C., and D. M. Gray. 1989b. CD measurements show that fd and IKe gene 5 proteins undergo minimal conformational changes upon binding to poly(rA). *Biochemistry.* 28:9502–9507.

Savitsky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36:1627–1639.

Shamoo, Y., A. M. Friedman, M. R. Parsons, W. H. Konigsberg, and T. A. Steitz. 1995. Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature.* 376:362–366.

Skinner, M. M., H. Zhang, D. H. Leschnitzer, Y. Guan, H. Bellamy, R. A. Sweet, C. W. Gray, R. N. H. Konings, A. H.-J. Wang, and T. C. Terwilliger. 1994. Structure of the gene V protein of bacteriophage f1 determined by multiwavelength x-ray diffraction on the selenomethionyl protein. *Proc. Natl. Acad. Sci. USA.* 91:2071–2075.

Stassen, A. P. M., G. J. R. Zaman, J. M. A. van Deursen, J. G. G. Schoenmakers, and R. N. H. Konings. 1992. Selection and characterization of randomly produced mutants of gene V protein of bacteriophage M13. *Eur. J. Biochem.* 204:1003–1014.

Terwilliger, T. C. 1996. Gene V protein dimerization and cooperativity of binding to poly(dA). *Biochemistry.* 35:16652–16664.

Thompson, T. M., B. L. Mark, C. W. Gray, T. C. Terwilliger, N. Sreerama, R. W. Woody, and D. M. Gray. 1998. Circular dichroism and electron microscopy of a core Y61F mutant of the f1 gene 5 single-stranded DNA binding protein and theoretical analysis of CD spectra of four Tyr → Phe substitutions. *Biochemistry.* 37:7463–7477.

Yu, L., C.-X. Zhu, Y.-C. Tse-Dinh, and S. W. Fesik. 1995. Solution structure of the C-terminal single-stranded DNA-binding domain of *Escherichia coli* topoisomerase I. *Biochemistry.* 34:7622–7628.

von Hippel, P. H., S. C. Kowalczykowski, N. Lonberg, J. W. Newport, G. D. Stormo, and L. Gold. 1982. Autoregulation of gene expression. Quantitative evaluation of the expression and function of the bacteriophage T4 gene 32 (single-stranded DNA binding) protein system. *J. Mol. Biol.* 162:795–818.

Walstrom, K. M., J. M. Dozono, and P. H. von Hippel. 1998. Effects of reaction conditions on RNA secondary structure and on the helicase activity of *Escherichia coli* transcription termination factor rho. *J. Mol. Biol.* 279:713–726.

Webster, R. E., R. A. Grant, and L. A. W. Hamilton. 1981. Orientation of the DNA in the filamentous bacteriophage F1. *J. Mol. Biol.* 152: 357–374.

Wold, M. S., and T. Kelly. 1988. Purification and characterization of replication protein A, a cellular protein required for *in vitro* replication of simian virus 40 DNA. *Proc. Natl. Acad. Sci. USA.* 85:2523–2527.

Zaman, G. J. R., A. M. Kann, J. G. G. Schoenmakers, and R. N. H. Konings. 1992. Gene V protein-mediated translational regulation of the synthesis of gene II protein of the filamentous bacteriophage M13: a dispensable function of the filamentous-phage genome. *J. Bacteriol.* 174:595–600.

Zaman, G., A. Smetsers, A. Kaan, J. Schoenmakers, and R. Konings. 1991. Regulation of expression of the genome of bacteriophage M13. Gene V protein regulated translation of the mRNAs encoded by genes I, III, V and X. *Biochim. Biophys. Acta.* 1089:183–192.

Zhang, H., M. M. Skinner, W. S. Sandberg, A. H.-J. Wang, and T. C. Terwilliger. 1996. Context dependence of mutational effects in a protein: the crystal structures of the V35I, I47V, and V35I/I47V gene V protein core mutants. *J. Mol. Biol.* 259:148–159.