PHYSICAL CHEMICAL STUDIES OF THE STRUCTURE AND FUNCTION OF DNA BINDING (HELIX-DESTABILIZING) PROTEINS*

Authors: **Joseph E. Coleman**
John L. Oakley
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

Referee: Peter H. von Hippel
Department of Chemistry
Institute of Molecular Biology
University of Oregon
Eugene, Oregon

I. INTRODUCTION**

Binding of proteins to DNA is fundamental to the mechanism of the control of gene expression in both prokaryotic and eukaryotic cells. Knowledge of the specific molecular features of DNA recognized by complementary features of the three-dimensional structure of the DNA-binding proteins is still in its infancy, but two large classes of DNA-binding proteins have emerged. One class includes the enzymes such as the polymerases and the restriction endonucleases, and the nonenzymatic repressor proteins which recognize specific sequences present in only one or a few copies per genome. A second group is made up of nonspecific DNA-binding proteins which form stoichiometric complexes with and modulate subsequent enzymatic transformations of the DNA. The histones, which form complexes with double-stranded DNA, may be regarded as the most prominent of the latter group.

Since the introduction of DNA-cellulose affinity chromatography by Alberts et al. in 1968,[1,2] a second group of proteins forming stoichiometric complexes with DNA regardless of sequence has emerged. Originally termed "unwinding proteins", these proteins have in common a higher affinity for single-stranded than for double-stranded DNA and thus promote the melting of double-stranded DNA. They are better termed helix-destabilizing proteins to distinguish them from the enzymes which "unwind" the helix by breaking and making phosphodiester bonds.[3]

Because the helix-destabilizing proteins form complexes with all single-stranded DNA regardless of base sequence, the molecular details of complex formation have been accessible to direct physicochemical measurements. Structural conclusions derived with techniques which include chemical modification, UV spectroscopy, circular dichroism, NMR, and X-ray diffraction will be reviewed. The following proteins will be discussed in detail: the gene 32 protein of bacteriophage T4, the gene 5 protein from bacteriophage fd, and the helix-destabilizing protein from *Escherchia coli*. The largest amount of structural information is available for the gene 5 protein and specific models

* Original work from the authors' laboratory was supported by NIH Grant GM 21919-04 and NSF Grant PCM76-82231

** Abbreviations used: CD, circular dichroism; NMR, nuclear magnetic resonance; HDP, helix-destabilizing protein; gene 32P, gene 32 protein of T4; gene 5P, gene 5 protein of fd (M13); gp32*1 or gene 32P*-A, gene 32 protein with the C-terminal "A" peptide removed; MSV, murine sarcoma virus; and RLV, Rauscher leukemia virus

for this protein, and its complexes with DNA based on NMR and X-ray diffraction data are presented. A number of other helix-destabilizing proteins from both prokaryotes and eukaryotes has been described, and a survey of these will be given.

Some of the basic molecular features of DNA-protein interactions emerging from studies of the helix-destabilizing proteins are likely to be shared by DNA-binding proteins like the RNA polymerases and repressors which recognize specific nucleotide sequences. Properties of some of these more complex systems which suggest this will be discussed.

II. GENE 32 PROTEIN OF BACTERIOPHAGE T4

A. Molecular Biology of Gene 32 Protein

Gene 32 protein (gene 32P), coded for by the bacteriophage T4 genome, was the first helix-destabilizing protein to be described in detail.[1-4] It was with the description of the structure and function of gene 32P that it was first shown that a single-stranded DNA-binding protein which acted in a stoichiometric rather than a catalytic fashion could be involved in enzymatically catalyzed transformations of DNA.[4-6] Normal replication of T4 DNA requires the functioning of 19 T4 genes.[7,8] While the products of many of these are not well characterized, these genes include gene 43, which codes for the T4 DNA polymerase,[9-11] and gene 32, which codes for a protein of mol wt 35,000.[11] The product of gene 32 was known to be required for replication,[7] repair,[12] and recombination[13] of T4 DNA. Two experimental examples illustrating this are as follows: an amber mutant in gene 32 requires 40 min at 37°C to complete even a single round of replication[14] (compared to several minutes for one round of DNA replication in normal T4 infection; total phage DNA is present between 10 and 20 min) and T4 mutant ts P7, temperature sensitive in the gene 32 product, which replicates normally at 25°C, while replication ceases in 1 min at 42°C.[15] The gene 32P is made both early and late in T4 infection and ~10,000 copies accumulate in the infected *E. coli* cell.[4]

Gene 32P binds strongly to single-stranded DNA regardless of sequence, and large quantities of the protein can be isolated by stepwise elution from a single-stranded DNA-cellulose column.[4] Salt at concentrations of 0.6 *M* or greater is required for complete elution, suggesting a strong electrostatic component to the binding, even though the native protein carries a net negative charge at neutral pH[4] (see below for discussion of alteration in charge on limited proteolysis). Gene 32P completely covers a circular single-standed DNA like fd when added to the pure DNA in vitro. A diagram of the gene 32P-fd DNA complex is shown in Figure 1. In contrast to the helix-destabilizing protein from fd, gene 5P, gene 32P does not collapse the DNA. Alberts and Frey[4] showed that binding of gene 32P to fd DNA was highly cooperative. Different amounts of gene 32P (labeled with the same amount of tritiated gene 32P) were each mixed with 10 μg (a large excess) of fd DNA at an elevated salt concentration where the complex is marginally stable, and the tritiated complex detected by sucrose density centrifugation. A 14-fold increase in gene 32P concentration resulted in a dramatic increase in protein-DNA binding, from no detectable binding at 0.5 μg of protein to 60 to 70% binding at 7 μg of protein. In order to explain the dramatic increase, it was estimated that the affinity of a gene 32P monomer for a site adjacent to an already bound monomer (a contiguous site) must be at least 80 times greater than for an isolated site on the DNA. This property is probably due to side-side interactions between protein molecules, and such interactions can indeed be seen in vitro in the absence of DNA. Self-association has been demonstrated on sucrose gradients[6] by equilibrium sedimentation and on polyacrylamide gels,[16] and by analytical ultracentrifugation.[17] At moderate protein concentrations and high salt concentrations, gene 32P is present as a strongly associated dimer. On lowering the ionic strength, larger aggregates ap-

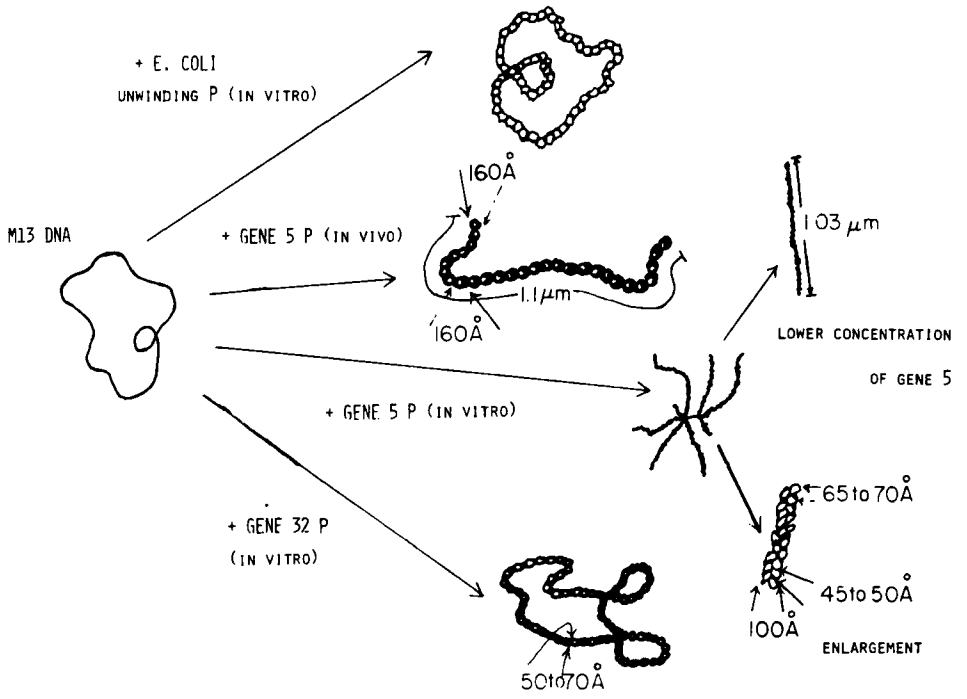


FIGURE 1. Schematic representation of the appearance of the complexes of fd (M13) DNA with three helix-destabilizing proteins, gene 32P of T4, gene 5P of M13 (both in vitro and in vivo complexes shown) and *E. coli* helix-destabilizing protein. Dimensions given are from glutaraldehyde cross-linked preparations shadowed directly without mounting on protein films to give as little distortion of dimensions as possible.

pear, and these are suggested to be linear aggregates with their DNA binding grooves aligned, the interactions between them being those responsible for the cooperative nature of the binding to DNA.[16,17] The bonds between gene 32P monomers in the discrete dimer form are clearly of a different kind than those responsible for the higher aggregates, since they are stable under conditions not conducive to high aggregation.[16,17] Carroll et al.[16] have incorporated both kinds of interaction into two possible models of gene 32P action at a replication fork, although it is difficult to reconcile their suggestions with the fact that gene 32P complexes with single-stranded circular DNAs are not collapsed into a rod as with gene 5P complexes (see below). It remains uncertain, therefore, whether the intersubunit bonds of the dimer form remain intact in the DNA complex or are of any physiological significance.[18]

The presence of gene 32P lowers the melting temperature of double helical DNAs.[19] The T_m of poly[d(A-T)] is lowered from 65 to 25°C.[19] The ease of melting of double-stranded DNAs may well relate to the A·T base pair content of the DNA, but other factors may also be involved. Some data suggest that certain double-stranded DNAs may have kinetic rather than equilibrium blocks to melting. The ease of melting of poly[d(A-T)] may result from looped configurations of this polymer which allow nucleation points for melting processing.[24,29] It is somewhat surprising to find that native gene 32P does not melt T4 DNA at 37°C, or at least not major regions of it.[4] On the other hand, there is electron microscopic evidence that a proteolytically modified gene 32P (minus the ~50 C-terminal amino acid residues) does melt regions of the T4 genome.[20] In the last several years, the nature of the modification of the gene 32P by limited proteolysis has been worked out in detail[21,22] and will be discussed below.

Although gene 32P melts double-stranded DNA, it also catalyzes the renaturation of denatured DNA.[4] As indicated by a sixfold increase in the frictional coefficient of fd DNA when gene 32P binds, the DNA is held in a highly expanded form by the protein.[4] This expansion can also be observed directly by electron microscopy which shows the complex to have a greater contour-length, ~3 μm, than fd DNA alone, ~1.9 μm.[19] This suggests that gene 32P aids DNA renaturation by holding the DNA in a more uniform expanded form, lowering the kinetic barrier to reformation of the double helix under equilibrium conditions which favor duplex formation rather than the single-stranded DNA-gene 32P complex.

Gene 32P stimulates the rate of DNA synthesis catalyzed by the T4 DNA polymerase by five- to tenfold in a purified in vitro system.[6] While the maximally stimulated rate is only ~10% of that measured for the movement of in vivo growing points,[6] the stimulation is likely to be physiologically important, since the gene 32P from the temperature-sensitive mutant, tsP7, is also temperature sensitive in the in vitro stimulation of polymerase action. Stimulation of the T4 DNA polymerase activity is maximal at a level of gene 32P sufficient to bind to all the template present. While these results suggest that gene 32P stimulates the polymerase by removing inhibitory secondary structure from the single-stranded template and holding it in an optimal conformation, a weak complex between gene 32P and the polymerase can be detected.[6]

The above findings have led Moise and Hosoda[23] to propose a model in which gene 32P functions to maintain the proper topological relationships at the replication fork during DNA synthesis. Not only may gene 32P do this by binding to the template DNA, but possibly also by inducing protein-protein interactions with other protein components of the replication apparatus (see discussion below, Section II.C.1).

B. Physical Chemistry of Gene 32 Protein Binding to DNA

The earliest studies monitored binding of gene 32P to DNA by following the hyperchromism of the base chromophores (ΔA_{260}) produced on melting of double-helical DNA (e.g., poly[d(A-T)] by gene 32P. It was shown that the binding affinity decreases substantially with increasing Na^+ concentrations (0.15 to 0.60 *M*).[4] Similar cation effects on binding affinity to single-stranded DNAs or on the melting of double-stranded DNA have been amply confirmed by this and other methods, particularly those using the circular dichroism change in the base chromophores accompanying binding as the assay of complex formation.[21,24] Divalent cations (Mg^{++}) are more effective than monovalent cations in reducing binding either to initially single-stranded DNA or the melting of double-stranded DNA.[21] These findings suggested a large electrostatic component to the binding forces between the protein and DNA, even though gene 32P has a net negative charge at neutral pH[25] (see below for possible modification of changes on limited proteolysis). The effect of elevated ionic strength is directly on the protein-DNA interaction, not just caused by stabilization of the double helix, since it inhibits single-stranded binding as well. A higher sensitivity to salt is seen in the melting reaction due to the cation stabilization of the double helix. Lysyl and arginyl side chains of the protein are likely to contribute the electrostatic component.

An electrostatic interaction might give rise to some interaction with the double-stranded polyphosphate backbone (unless severe stereochemical restraints are present) and Jensen et al.[24] have devised a microsedimentation velocity technique with which to measure the affinity of gene 32P for native DNA under nonmelting conditions. Association constants, K, of 0.8×10^4 *M* (0.05 *M* Na^+) to 3.9×10^4 *M* (0.02 *M* Na^+) were measured. There is a roughly linear relationship between the log of K and [Na^+], suggesting a significant electrostatic component. Binding to duplex DNA is not cooperative.[24] These constants should be compared to the estimated value of K between

gene 32P and single-stranded DNA of $>10^9$ *M*, as determined from sucrose density centrifugation.[4] The free energy for the latter binding must contain contributions in addition to those of the electrostatic component and the cooperative interaction between contiguous monomers.

Jensen et al.[24] have carried out careful measurements of the hyperchromism of the base chromophores on the binding of gene 32P. For a double-stranded polynucleotide, this hyperchromism is greater than for temperature-induced melting, and for single-stranded DNAs there is an additional hyperchromism induced by gene 32P. The most reasonable conclusion is that gene 32P holds the single strand in a reasonably extended, if not fully extended, form and that it removes any residual base stacking from the single strand. In support of this conclusion, gene 32P binding does not induce hyperchromism in poly(dT), the only single-stranded homopolymer lacking significant residual base stacking. Calculating from the contour length of the in vitro complex of the gene 32P with fd DNA, there would have to be at least 5.3 Å per base. This number is calculated using the recent determination of 5740 ± 210 bases in fd DNA[27] rather than the earlier estimate of ~6600 bases.[26] The latter number predicts 4.6 Å for the separation of bases in the gene 5-fd DNA complex. Such calculations give minimum separations, since supercoiling almost certainly occurs to some degree, and hence the expansion may even approach that of a fully extended chain, ~7 Å per base.

Gene 32P causes significant shifts in the near UV circular dichroism of DNAs and synthetic homopolymers. These changes are similar in all polynucleotides and can be generally described as a decrease in the ellipticity of the positive bands between 260 and 280 nm, although the lower wavelength bands (260 to 270 nm) tend to be more substantially affected, leading to a movement of the cross-over point near 250 nm to longer wavelength. Although some of these CD changes must reflect uncoupling of the base chromophores (breaking the coupling found in a stacked structure and responsible for hypochromia as well as some of the ellipticity), the CD changes are different from those found in heat-induced denaturation and must reflect a particular conformation induced by the protein. CD changes in the base chromophores also accompany the binding of gene 32P to poly(dT), showing that apparently normal complex formation occurs even though significant additional hyperchromism does not occur because of the initially extended form of this polymer.

Changes in the ellipticity of the base chromophores of a variety of nucleic acids are induced by all three of the well-characterized HDPs (gene 32P,[21,24,29,30] gene 5P[28] and *E. coli* unwinding protein[21]). Qualitatively the CD changes induced in the same polynucleotide by all three proteins are remarkably similar, although of considerably greater magnitude in the case of gene 5P (see below). On the other hand, the gross features of the DNA complexes with the three proteins as visualized by electron microscopy are clearly different (Figure 1). Despite these gross differences in the conformation of the complexes, there must be great similarities in the uncoupling effects at the base-base interaction level induced by the proteins. There is evidence in the case of both gene 32P and gene 5P that tyrosyl ring intercalation with the bases may occur.[21,28,31-33] A structure involving tyrosyl-base of tyrosyl-base-phenylalanyl intercalation might contribute significantly to the CD changes, but such a structure is supported by extensive data only in the case of gene 5 protein (see discussion below).

Both spectrophotometric titrations (using the hyperchromism induced by gene 32P)[14,24] and CD titrations (using the Δ ellipticity induced by gene 32P)[21] of polynucleotides have been used to measure the site size for gene 32P binding, i.e., the number of bases covered by the cooperative binding of each gene 32P monomer. The initial CD titrations indicated an end point at ~10 nucleotides per gene 32P monomer[21] in agreement with earlier studies which determined the stoichiometry by sucrose density gradient centrifugation of a fixed quantity of [^{3}H]-gene 32P to which varying amounts

of fd DNA were added.[4] More recent highly precise titrations following the increase in poly(dA) hyperchromicity have shown the increase in hyperchromicity to be complete at a nucleotide to protein ratio of ~6.7:1.[24] Site size estimated from the melting profiles of gene 32P-complexed poly[d(A-T)] as a function of varying gene 32P concentrations were ~7.5 ± 0.3 nucleotides per protein.[24] The reason for these discrepancies is not clear, although due to the relatively small amplitude of the CD changes, the precision of the CD titrations is not as satisfactory as in the case of gene 5P where these changes are much larger.[28]

Melting curves for poly[d(A-T)] in the presence of varying amounts of gene 32P consist of two melting regions (one near 14°C reflecting complex formation and one near 41°C reflecting melting of uncomplexed poly[d(AT)]), separated by a plateau region. Such curves can be simulated rather precisely by choosing values for site size, n_c, the binding constant without cooperativity, K_c, and the cooperativity parameter, ω_c (defined as the equilibrium constant for the translocation of a binding protein molecule from an isolated to a contiguous binding site).[24] The observed affinity constant would be $K_c\omega_c$. The best fits were obtained setting $n_c \cong 7.5$ nucleotides per protein, $K_c \cong 10^7\ M^{-1}$ and $\omega_c \cong 10^3$. K_c is very sensitive to ionic strength, while ω_c is relatively ionic-strength independent. The physicochemistry of gene 32P binding to DNA has been recently reviewed by von Hippel et al.[29]

1. Intrinsic Chromophores of Gene 32 Protein

The optical activity of the peptide bond chromophores of gene 32P give rise to ellipticity bands typical of globular proteins. Mean residue ellipticities at representative wavelengths are $[\theta_m]_{208} = -9900$, $[\theta_m]_{217} = -7650$, $[\theta_m]_{22} = -6050$.[21] Application of the graphical method of Greenfield and Fasman using these values predicts a secondary conformation containing ~20% α-helix, ~20% β-structure, and 60% random coil.[21] A recent analysis based on a method devised by Yang reports values of 21.6% α-helix, 25.8% β-pleated sheet, and 52.6% random coil.[29,30] In addition to the intrinsic optical activity of the peptide bond chromophores, there is a number of unusually large ellipticity bands in the region of the near UV absorption bands of the tyrosyl and tryptophanyl side chains.[21] Whether these reflect any specific structure involved in nucleotide binding is unclear as yet, but both tryptophanyl and tyrosyl residues appear to be involved in DNA binding (see below).

The intrinsic fluorescence of gene 32P is dominated by the fluorescence of tryptophanyl side chains.[18] The emission maximum is at 347 nm when the protein is excited at 290 nm. Tyrosyl fluorescence, if any, must involve efficient energy transfer to the indole ring, since excitation at 270 nm produces no difference in emission. Large oligo- and polydeoxynucleotides to which gene 32P binds cooperatively quench the fluorescence by 20 to 35%.[18] In contrast, most short nucleotides to which cooperative binding is not possible quench only 2 to 6%. The exceptions are those like $d(Ap)_2$ (but not d(ApA) or $d(pA)_2$), which contain the 3′ phosphate and produce ~17% quenching. This has suggested that the monoester phosphate may quench a tryptophan near the DNA-binding groove. Cooperative binding may also alter conformation of the protein and induce a change in the environment of one or more tryptophanyl side chains. In view of the possible participation of tyrosyl side chains in DNA binding, a fluorescence mechanism involving energy transfer from tyrosyl to tryptophanyl side chains interrupted by nucleotide binding (possibly intercalation) should not be ruled out.

Kelly et al.[34] have used the fluorescence quenching to measure the site size of the gene 32P nucleotide binding surface and find that the fluorescence is totally quenched at a nucleotide to protein ratio of ~5:1, which is lower than the numbers found by

the other methods. Thus, site size for gene 32P is somewhat uncertain within the range five to ten nucleotides per protein monomer.

C. Enzymatic and Chemical Modification of Gene 32 Proteins

1. Limited Proteolysis of Gene 32 Protein; Relationships to DNA Binding

Gene 32P, even in relatively highly purified form, has been observed to undergo on storage a proteolysis which reduces the molecular weight to < 30,000.[21] A similar change is induced rapidly by trypsin[20,21,23] and extracts of T4-infected cells are known to have tryptic activity.[35] Preliminary studies suggested that the cleavage by trypsin was highly specific and removed ~50 amino acid residues from the C-terminal end of the protein, a region containing 17 of the 35 Asn residues of the protein and therefore likely to be negatively charged.[21] The general structure of the protein as measured by circular dichroism is not altered by this cleavage, and the modified protein binds as tightly if not more tightly to DNA.[21] Hosoda et al.[20] have shown that the C-terminally cleaved protein binds more tightly to DNA, since it appears to melt T4 DNA at physiological temperatures and will even bind to double-stranded DNA-cellulose (possibly because it melts certain regions). The tighter binding of gene 32P from which the C-terminal region has been removed (gp32*1, mol wt 27,000) has been confirmed in quantitative fashion by using the native and modified proteins to melt poly[d(A-T)] as followed spectropolarimetrically. gp32*1 lowers the T_m by ~12°C more than the native protein.[30] The conformation of the DNA is the final single-stranded complex appears to be similar.[30]

While the limited proteolysis may or may not be a feature of the in vivo cycle of gene 32P, some interesting structural alterations in gene 32P on binding to DNA have been detected recently using its susceptibility to tryptic hydrolysis as the probe. In connection with the studies of the primary structure of gene 32P, Williams and Koigsberg[22] have defined two regions of the protein, N- and C-terminal polypeptides, which can be removed by trypsin. Similar tryptic cleavages were described by Moise and Hosoda,[23] although no structural work was carried out. The C-terminal contains 59 amino acids and is known as the A region, while the N-terminal region is known as the B region, and present data suggest it to contain 28 amino acids. The general features are outlined in Figure 2. The sequence of the B region is Met-Phe-Lys-Arg-Lys-Ala-Thr-Ala-Glu-Leu-Ala-Ala-Gln-Met-Ala-Lys^{22,36}-[Leu-Asn-Gly-Asn-Lys-Gly-Phe-Ser-· · · -Glx-Asn-Lys][22] as determined by both Williams and Konigsberg[22] and Tsugita and Hosoda.[36] The amino acid composition of the A polypeptide is given in Table 1 along with the composition of the native gene 32P.[21,22]

Poly(dT), poly(dC), and single-stranded DNA increase the rate at which trypsin removes the C-terminal A region, but these polynucleotides decrease the susceptibility of the B region to tryptic hydrolysis.[22] In contrast, oligonucleotides which are too short to allow cooperative binding of gene 32P have no effect on the rate of removal of A and B by trypsin. The fact that the A and B peptides are relatively susceptible to tryptic hydrolysis, while the core molecule is not, suggests that the N- and C-terminal regions must form domains on the surface, perhaps relatively flexible and extended. There is some NMR evidence for this (Section II.D). The opposite changes in susceptibility to hydrolysis of the A and B regions upon binding to DNA led Williams and Konigsberg[22] to propose a model for cooperative binding of gene 32P to the polynucleotide chain as pictured in Figure 3. The flexible region B is suggested to be involved in the side-to-side monomer-monomer interactions; hence its occlusion in the complex. Removal of the B region results in a decrease in DNA binding affinity.[22] On the other hand, the A flexible region is pictured as swinging out, possibly in a position to interact with an-

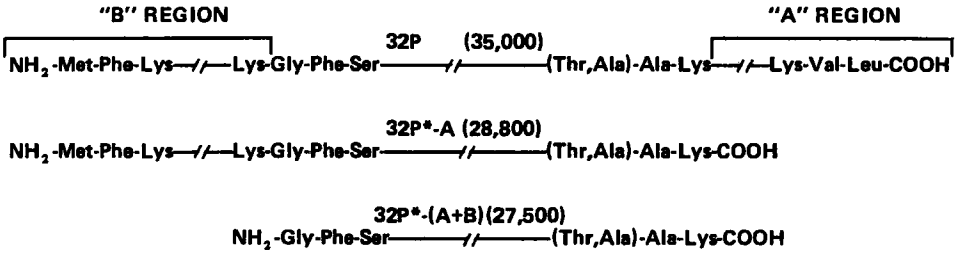


FIGURE 2. Summary of the partial amino acid sequence data on gene 32P, gene 32P*-A, and gene 32P*-(A + B).[22] (From Williams, K. R. and Konigberg, W., *J. Biol. Chem.*, 253, 2463, 1978.)

TABLE 1

Amino Acid Composition of Gene 32P and Its A Peptide

Amino acid	Gene 32P	A
Cysteine	4.6	0
Aspartic acid	53.6	17.1
Threonine	14.0	3.4
Serine	24.3	7.0
Glutamic acid	32.2	2.6
Proline	10.5	0.4
Glycine	21.1	1.8
Alanine	27.0	4.9
Valine	20.8	3.6
Methionine	9.8	2.9
Isoleucine	10.8	1.1
Leucine	21.5	6.8
Tyrosine	8.5	0.6
Phenylalanine	18.6	3.4
Histidine	2.5	0
Lysine	35.1	3.2
Arginine	4.5	0.5
Tryptophan	5.5	0
TOTAL	325	59

other protein of the replication complex. A possible candidate for such an interaction is the gene 44P-gene 62P complex, which is also required for T4 DNA replication and will bind to a single-stranded DNA-gene 32P complex, but not to gene 32P or single-stranded DNA alone.[37]

2. Chemical Modification of Gene 32 Protein — Effect on DNA Binding

Chemical modifications of gene 32P have not been as extensive as those of gene 5P, which will be discussed below, but the consequences of nitration of tyrosyl residues are remarkably similar in both proteins. Treatment of the protein with tetranitromethane nitrates five of the nine tyrosyl residues in the protein and abolishes the tight binding of DNA.[21] The presence of fd DNA prior to adding the reagent completely prevents nitration of the tyrosyl residues. Evidence for a model of nucleotide binding

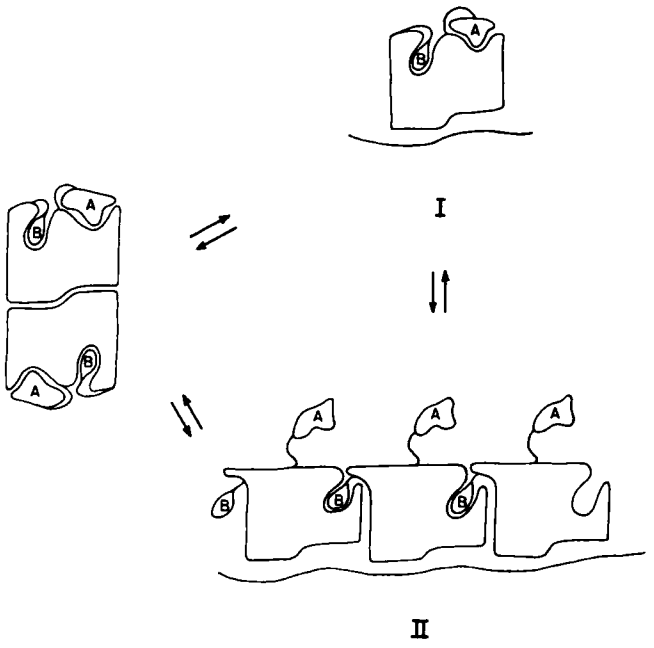


FIGURE 3. A model for the interaction of the gene 32P with oligonucleotides and single stranded DNA. The model assumes that gene 32P binds oligonucleotides (<8) without undergoing a change in conformation with respect to the A and B regions (Conformation I above). Contiguous binding to single-stranded DNA, however, induces a change from Conformation I to II. Only the latter conformation is suitable for cooperative protein-protein interactions.

to the gene 5P, which involves intercalation of tyrosyl and phenylalanyl rings between the base rings, will be outlined below. It it possible that gene 32P-DNA binding shares a similar mechanism as part of the binding interaction.

D. NMR of Gene 32 Protein

In view of the fact that, at high protein concentrations, gene 32P forms long linear aggregates, it would be predicted that high resolution NMR studies requiring millimolar concentrations of protein would not give high resolution because of broadening due to the extremely long rotational correlation time of such aggregates. The ^{1}H NMR of gene 32P shows, however, a number of relatively sharp resonances that represent only a relatively small number of the residues present in the protein (Figure 4). In the lowfield region of the spectrum where the aromatic protons resonate, only resonances from phenylalanyl protons are present; none of the tyrosyl or tryptophanyl protons are observed. In the upfield region, a number of aliphatic methyl and $-CH-_2$ protons are observed, along with a significant number of lysyl ε-CH_2 and a few arginyl γ-CH_2 resonances.[137] The A and B peptides together contain six Phe, eight Lys, two Arg, and a large number of aliphatic residues, but none of the tyrosyl or tryptophanyl residues. Thus, one interpretation of the ^{1}H NMR spectrum is that the A and B regions have a good deal of rotational motion independent of the core protein, and hence the residues in these "tails" give rise to reasonably sharp resonances. On the other hand, the resonances from residues in the aggregated core, e.g., the nine Tyr, six Trp, and three His, are broadened beyond detection. The resonances observed are compatible with such a postulate. A similar division of structural domains appears to occur in the histone core

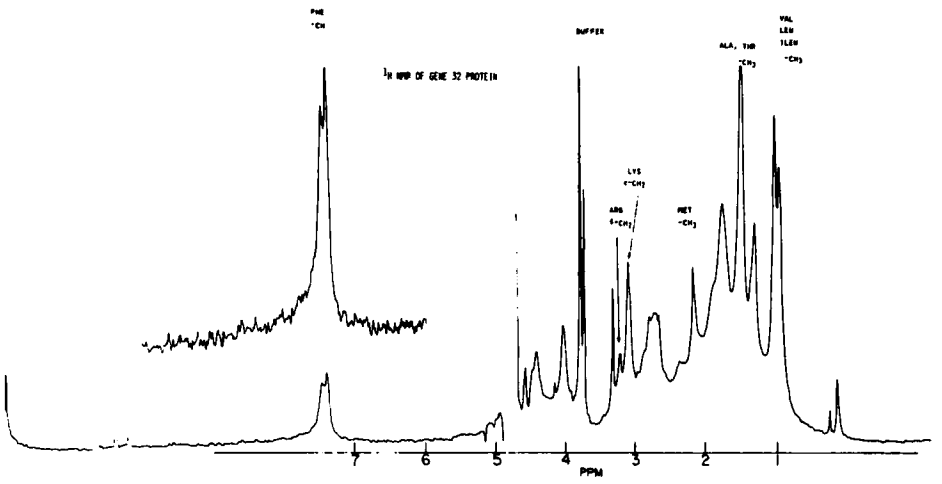


FIGURE 4. ^{1}H NMR spectrum of gene 32P (1×10^{-3} *M*) at pH 8, 25°C. (From Coleman, J. E., unpublished data.)

complex as first suggested by ^{1}H and ^{13}C NMR studies which showed a number of sharp resonances from this large complex.[38,39]

III. GENE 5 PROTEIN OF BACTERIOPHAGE fd

Gene 5P from the filamentous bacteriophage fd is the helix destabilizing protein for which the most detailed structural information is available. Phages of the Ff group (fd, M13, f1, and ZJ/2) are indistinguishable with respect to the properties discussed in this review. The features of the protein structure involved in interaction with single-stranded DNA have been identified more completely than for any other member of this class of proteins. Gene 5P contains 87 amino acids for a total molecular weight of 9689.[40-42] The complete amino acid sequence as determined by Nakashima et al. is given in Figure 5. Application of the Chou-Fasman analysis for predicting secondary structure by assigning a probability to each residue of occurrence in α-helix or β-structure (as determined from known secondary structures of crystalline proteins) predicts the predominant secondary structure favored by this sequence to be β-pleated sheet (Figure 5). Little or no α-helix is predicted, and the N- and C-terminal regions are not well predicted.[28] That β-sheet is the predominant secondary structure of the native molecule is confirmed by the crystal structure (see below). While there is general agreement between the prediction and the crystal structure on the presence of β-structure, the initial Chou-Fasman analysis and later refinements do not predict β-sheet in the region of residues 10 to 24. This appears from the crystal sructure to form the first strand of a three-stranded antiparallel β-sheet structure involving residues from about 10 to 41.[43] The relationship of this three-stranded structure to the DNA binding surface will be described below. A refined Chou-Fasman analysis[44,137] predicts β-turns at residues 15 to 18, 24 to 27, and 38 to 41, which are the β-turns defining this three-stranded region. Tyrosyl-41, which completes this region, is one of the residues believed to define the end of the DNA-binding surface according to present models (see below). The molecular biology of gene 5P is summarized briefly below before giving a detailed account of the solution and crystal structure data from which the present model of the gene 5P-DNA complex is derived.

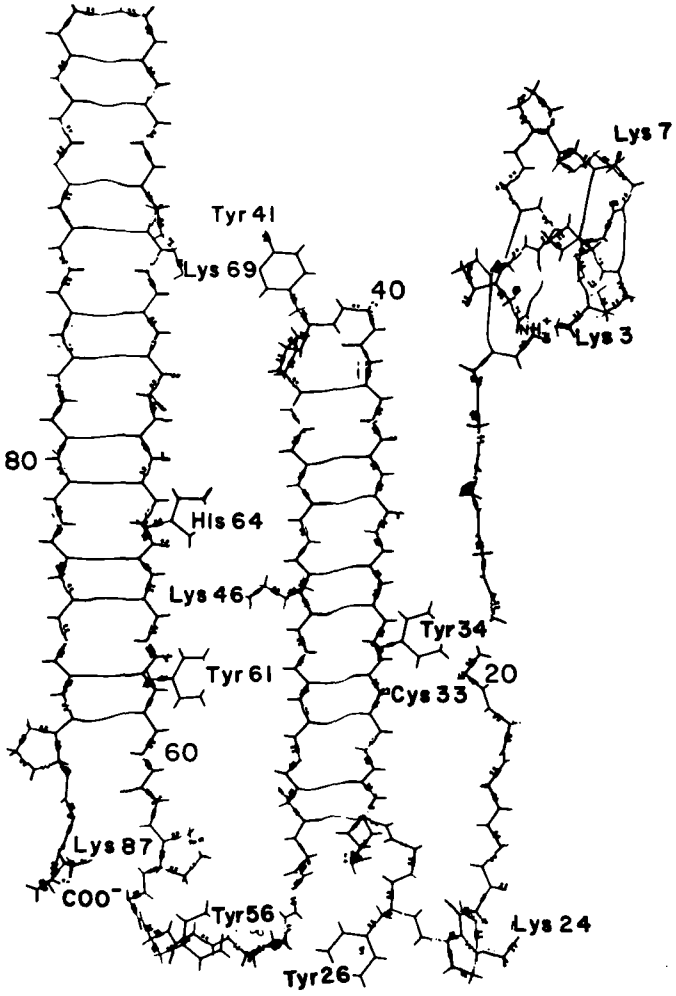


FIGURE 5. Secondary structure of gene 5P as predicted from the conformational parameters for amino acids in α-helical, β-sheet, and random coil regions as determined by the method of Chou and Fasman.[138] The sequence of gene 5P[40-42] is given below and the three tyrosyl residues modified by tetranitromethane are underlined. Met-Ileu-Lys-Val-Glu-Ileu-Lys-Pro-Ser-Gln[10]-Ala-Gln-Phe-Thr-Thr-Arg-Ser-Gly-Val-Ser[20]-Arg-Gln-Gly-Lys-Pro-Tyr-Ser-Leu-Asn-Glu[30]-Gln-Leu-Cys-Tyr-Val-Asp-Leu-Gly-Asp-Gln[40]-Tyr-Pro-Val-Leu-Val-Lys-Ileu-Thr-Leu-Asp[50]-Glu-Gly-Gln-Pro-Ala-Tyr-Ala-Pro-Gly-Leu[60]-Tyr-Thr-Val-His-Leu-Ser-Ser-Phe-Lys-Val[70]-Gly-Gln-Phe-Gly-Ser-Leu-Met-Ileu-Asp-Arg[80]-Leu-Arg-Leu-Val-Pro-Ala-Lys.[28] (Reprinted with permission from Anderson, R. A., Nakashima, Y., and Coleman, J. E., *Biochemistry*, 14, 907, 1975. Copyright by the American Chemical Society.)

A. Molecular Biology of Gene 5 Protein

The infectious cycle of fd (M13) coliphage is illustrated schematically in Figure 6. Only the main features are illustrated, and more extensive review articles on this subject should be consulted for details.[26,45] The filamentous phage attaches to the membrane of *E. coli* by a poorly understood mechanism which may involve the F pilus, as suggested by electron micrographs.[45] The attachment protein (product of phage gene 3) may be involved, and three copies of this protein exist at one end of each virion.

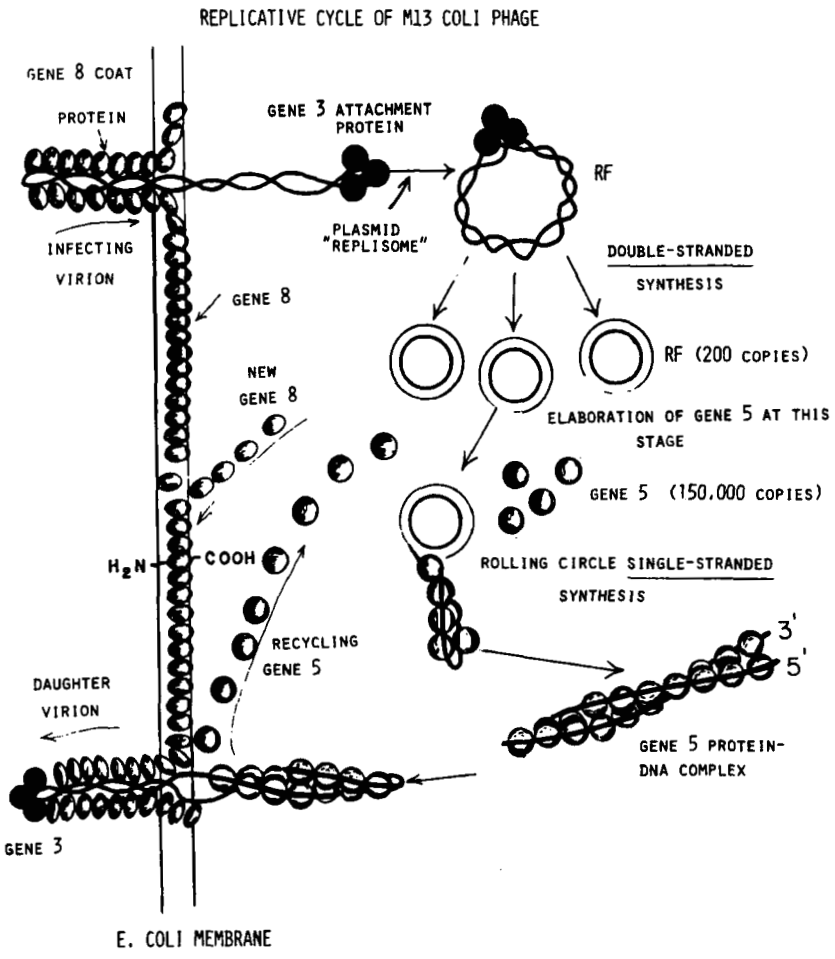


FIGURE 6. Diagram of the replicative cycle of M13 coliphage.

The rest of the protein of the intact phage is the coat protein of mol wt 5000 (product of phage gene 8). The sequence of gene 8P has been determined[46] and extensive studies of the structure of the protein and DNA in the whole phage have been made.[47,48] Circular dichroism of the whole phage as well as of gene 8P alone, solubilized, and dissolved in artificial membrane vesicles, suggests that the majority of the polypeptide backbone is in the α-helical conformation.[49] Application of a Chou-Fasman analysis to the sequence of gene 8P predicts that 80 to 90% of the residues will assume an α-helical structure.[137] Circular dichroism of gene 8P in lipid vesicles suggests a somewhat smaller percentage of α-helix.[49] Fiber X-ray diffraction patterns of whole phage do not reveal strong reflections from the DNA, hence the structure of the DNA in the phage is not known.[50-52] A "fish scale" model of the packing of gene 8P on the surface of the DNA has been proposed based on available sequence and structural data.[47,48]

As the DNA of the virion enters the bacterial membrane, gene 8P dissociates in the membrane. Tracer studies using radioactively labeled gene 8P show it to be reincorporated, diluted with newly synthesized protein, into the daughter virions. The surface of the phage, composed of the exposed portions of the gene 8P molecules, is highly water soluble, yet the dissociated gene 8P is highly insoluble in water and can only be solubilized by detergents.[53] It is soluble in lipid vesicles.[49]

Gene 8P is initially synthesized as a larger polypeptide chain containing an N-termi-

nal extension of 23 residues of a largely hydrophobic nature.[54] A preparation of isolated inverted membrane vesicles from uninfected *E. coli* has been shown to catalyze the specific removal of this signal peptide.[54] Thus it would appear that the synthesis and membrane-insertion of gene 8P follows the now classic "signal peptide" hypothesis initially proposed for eukaryoic secretory proteins in which the hydrophobic N-terminal extension or signal peptide is responsible for insertion and orientation of the protein in the membrane.[55]

Wickner[56] has shown that the insertion of gene 8P into the *E. coli* cytoplasmic membrane is asymmetric. Antibodies prepared against the whole virion react with the N-terminal eight amino acid residues of gene 8P. At all stages of infection by M13, this antigenic site is exposed to the cell exterior. The amino terminus of the coat protein is also found exclusively on the outside of dilauroyl or dimyristoyl lecithin vesicles containing gene 8P. Of the eight N-terminal residues, 95% were released from the vesicles by tryptic hydrolysis at the Lys 8-Ala 9 bond.[56] Likewise, chymotrypsin released the N-terminal region by hydrolysis of the Phe 11-Asp 12 bond. On the other hand, chymotrypsin trapped inside the vesicles released the C-terminal region by hydrolysis of the Trp 26-Ala 27 bond.[56] This leads to a model in which gene 8P extends through the membrane with the N-terminal acidic portion on the outside and the C-terminal basic region exposed on the inside, while the middle hydrophobic portion of the polypeptide is within the membrane. The carboxyl terminal section possibly reacts with the DNA polyanion as it enters the cytoplasmic surface of the membrane from the gene 5P complex. The NH_3^+-terminus may end up on the outside of the virion in a position to react with solvent. The outside position of the N terminus is supported by the fact that the antibodies against the whole virion react with the N terminus. This supports an earlier model of Marvin and Wachtel,[47] who proposed that the C terminus was on the interior of the cell membrane and reacted with the DNA polyanion to aid its passage through the membrane.

Once the DNA is in the cytoplasm, there is a period of DNA synthesis producing ~200 copies of the double-stranded RF form of the phage.[26] Double-stranded synthesis is catalyzed by the "plasmid replisome" (see Kornberg[45] for details). In vitro DNA synthesis systems suggest that the *E. coli* unwinding protein may play an activating role in RF synthesis[45] (see Section IV). Following this step, DNA synthesis changes to predominantly synthesis of viral single strands. Gene 5P is required for the switch from double-stranded to single-stranded DNA synthesis. Infection of *E. coli* with mutants of fd temperature sensitive in the gene 5 product results in normal synthesis of daughter virions at the permissive temperature, but failure to initiate single-stranded synthesis at the nonpermissive temperature.[57,58] As a consequence, double-stranded forms accumulate.

Electron micrographs of osmotically shocked spheroplasts prepared from fd-infected *E. coli* showed the presence of long cigar-shaped complexes between protein and DNA which did not contain gene 8P.[59] The only protein associated with the DNA (which was the circular single-stranded DNA of the daughter virions) was shown to be gene 5P.[59] The dimensions and contours of this in vivo gene 5P-fd DNA complex are indicated schematically in Figure 1. Thus, the major function of gene 5P appears to be to bind to the emerging single-stranded daughter virion, as it is synthesized on the complementary strand of the nicked RFs, and prevent its use by the host enzymes for synthesis of the complementary strand. Addition of gene 5P (in amounts stoichiometric to the DNA) to an in vitro reconstituted RF-synthesizing system halts DNA synthesis.[60] There is a suggestion that there may be an additional positive role for gene 5P in initiating single-stranded viral DNA synthesis.[61] The major function of gene 5P is indicated in Figure 6 by showing gene 5P complexing the single-strands emerging

from a "rolling circle" DNA synthesis. The in vivo complex apparently migrates to the cytoplasmic surface of the membrane where the DNA passes through the membrane, collecting its complement of gene 8P and gene 3P in the process. Gene 5P remains in the cytoplasm. Infected *E. coli* build up a content of $\sim 10^5$ molecules of gene 5P per cell, which must represent an equilibrium value for this continuously replicating phage.[62] Gene 5P is a relatively hydrophobic protein,[28] but soluble in low salt up to 10 to 15 mg/mℓ. The mechanism involved in removing gene 5P and packaging the DNA in gene 8P as the DNA polyanion moves through the membrane is unknown at present, but its physicochemistry is intriguing (see discussion of crystal structure).

B. Physical Chemistry of Gene 5 Protein Binding to DNA

The in vivo gene 5 P-fd DNA complex as visualized by electron microscopy is approximately 1.1 μm long × 160 Å wide (Figure 1).[59] Since the DNA is a covalent circular single strand of about 5700 bases,[27] gene 5P must collapse the circle into two antiparallel strands within at least 100 Å of each other. Careful inspection of the high-power electron micrographs suggest that some sort of supercoiled structure is present and accounts for the observed 160-Å width. Even if the minimum of twice the complex length, 2.2 μm, were taken as the length of the DNA, the base-to-base distance would be ~3.8 Å per base.[21] Since the DNA almost certainly follows some sort of superhelical path, the base-to-base distance in the complex is probably considerably greater, and the DNA chains may be almost fully extended as present models of the gene 5P binding to a tetranucleotide suggest (see below).

If excess gene 5P is added to fd DNA in vitro, the gene 5P-DNA complex appears in electron micrographs as a multipronged structure of rods radiating from a central region (Figure 1).[62] This structure apparently arises from the cooperative binding property of gene 5P, i.e., the gene 5P-DNA complex nucleates by the binding of a gene 5P dimer at several points of the fd DNA circle. Subsequent binding of gene 5P dimers is then much more favorable at contiguous sites, collapsing the circle from several points to end up as a cloverleaf arrangement. As the gene 5P to fd DNA ratio is reduced, rod-shaped structures more closely approximating the in vivo complex appear (Figure 1). Dimensions of the in vitro complex as estimated from the shadowed electron micrographs appear to be significantly different from those of the in vivo complex. For example, the width appears closer to 100 than 160 Å. The origin of these apparent differences is not clear. The possibility that additional proteins are present in the in vivo complex has not been ruled out.

With the introduction of DNA-cellulose chromatography, both Alberts et al.[62] and Oey and Knippers[66] isolated large amounts of gene 5P by binding it to single-stranded DNA-cellulose columns followed by salt elution.

At least two studies of the hydrodynamic properties of gene 5P suggest that the homogeneous protein exists primarily as the dimer.[63,64] Cross-linking of the unliganded protein with suberimidate shows the formation of a cross-linked dimer at about 50% yield, also suggesting that the dimer is a major species in solution.[65] As will be detailed extensively below, a variety of experimental techniques show each gene 5P monomer to bind to about four nucleotide bases. NMR studies with isolated tetradeoxynucleotides also show tetranucleotide binding to mimic most of the specific interactions induced by longer deoxynucleotides.[31,32] Thus these findings must be accommodated by models for the topology of the extended complexes with whole fd DNA.

Equilibrium ultracentrifugation shows that addition of nucleotides induces the formation of higher oligomers of gene 5P.[64] Cross-linking of gene 5P in the presence of oligonucleotides results in cross-linked oligomers containing up to eight monomers, suggesting that the nucleotides induce oligomerization.[65] These studies suggest that

nucleotide binding induces additional conformational changes in the protein which lead to additional protein-protein interactions, compatible with the observation of cooperative binding.

It now appears that all of these findings can be accommodated by a model of complex formation in which a symmetrical dimer is the basic structural unit of gene 5P and in which the DNA is found on the outside of this dimer unit as illustrated in primitive fashion in Figure 6. The protein-protein interactions of the dimer hold the two antiparallel DNA strands together, although the two oppositely directed nucleotide binding sites are separated. Nucleotide interaction may stimulate additional side-to-side interactions of these dimers. The crystal structure of the unliganded protein shows a tightly interlocking dimer related by a twofold axis of symmetry. Additional structural features leading to the supercoiled structure 160 Å wide are speculative. One model, based on the finding that crystallization of gene 5P in the presence of oligonucleotides leads to a crystal containing a hexamer as the basic unit, will be shown below.

1. Optical Spectroscopy of Gene 5P and its DNA Complexes

More optical spectroscopy has been done on gene 5P than any other member of the HDP class, and the results have significant implications for its structure and function. The UV absorption spectrum has a maximum at 276 nm with a rapid fall off in the 290-nm region, reflecting the absence of tryptophan.[67] The near UV chromophores are the three Phe and five Tyr residues.[40,42] The most striking optical property of gene 5P is the presence of two large positive ellipticity bands in the circular dichroic spectrum with maxima at 228 nm ($[\theta] = 3.2 \times 10^5$ deg cm^2/dmol) and 200 nm ($[\theta] = \sim 8 \times 10^5$ deg cm^2/dmol).[28,67] Both bands are probably due to disymmetric perturbations of the tyrosyl chromophores and must reflect an unusual arrangement of these side chains, possibly the tyrosyl-phenylalanyl-tyrosyl stacking of the surface residues proposed for the DNA-binding groove (see below). Ellipticity of the band at 228 nm falls by ~30% with the binding of deoxynucleotides. While analogous model systems do not yet exist, it is possible that intercalation of bases between the aromatic side chains decreases the rotatory strength of the originally stacked chromophores.

The UV ellipticity bands are a product of the native conformation, since they disappear on denaturation. The two positive ellipticity bands completely dominate any contribution to the circular dichroism by the peptide chromophores, suggesting that the latter do not have large molar ellipticity, a finding favoring β or random conformations for these bonds, since the α-helix has large negative rotation at 222 nm, which should be easily observed even in the presence of the 228-nm band.

As in the case of gene 32P mentioned above, gene 5P induces hyperchromism in the absorption bands of the bases of single-stranded DNA, suggesting also that the phosphodiester backbone is held in an extended conformation.[62] Binding of gene 5P to both polydeoxynucleotides and polyribonucleotides results in dramatic changes in the ellipticity of the base chromophores.[28] For both poly[d(A-T)] and fd DNA, there is a dramatic reversal of the sign of the ellipticity near 270 nm from positive to negative. Although qualitatively similar to ellipticity changes observed on formation of complexes between gene 32P or *E. coli* HDP and these same polynucleotides, the magnitude of the change is much larger in the case of the gene 5P complex. This may relate to the collapsed antiparallel chain structure induced by gene 5P, in contrast to the circular single-stranded structure of the other two complexes (Figure 1). While these ellipticity changes undoubtedly reflect alteration in base stacking (possibly tyrosyl intercalation), extension of the backbone and a change toward the gauche-gauche conformation (see ^{31}P NMR results, Section III.D), it is difficult to interpret the circular dichroism of polynucleotides in terms of precise conformation. Removal of ordered

water molecules around the nucleotide, likely to accompany binding, may contribute to changes in ellipticity, as will vicinal effects from the protein.

The large magnitude of the change in ellipticity of the base chromophores in the gene 5P-nucleotide complexes has provided a more sensitive and convenient assay of complex formation for this protein than for any of the others. It has been used to follow the effects of chemical modification (see below). For oligonucleotides with lower binding constants than native single-stranded DNA, the concentrations employed in the CD assay can be lowered to the point where satisfactory determinations of the binding constants can be made by spectropolarimetric titrations. Several examples are given in Figure 7, plotting $[\theta]$ against the ratio 1/R; R = nucleotide to protein ratio. From these data and the concentrations of nucleotide and protein employed, it can be calculated that the sharp end point (for fd DNA or $d(pA)_n$) means a dissociation constant of $<10^{-8}$ *M*. $d(pT)_8$ and $d(pA)_4$ do not saturate at R values of 4:1, and binding constants of $\sim 3 \times 10^{-6}$ *M* and $\sim 10^{-7}$ *M* can be calculated. This was the first evidence that gene 5P, although binding to tetranucleotides (and presumably DNA) of any base composition and sequence, has a pronounced preference for A vs. T residues.[31] The possible physiological significance of this is unclear, although it might influence the location of initiation sites for the cooperative binding of gene 5P to DNA.

C. Chemical Modification of Gene 5 Protein — Effect on DNA Binding

A number of chemical modifications of the protein which interfere with deoxynucleotide binding or which are prevented by prior binding of nucleotides to the protein has been carried out in an attempt to identify the amino acid side chains involved in nucleotide binding.[28] A summary of the chemical modification data is given in Table 2. The assay for DNA complex formation with the modified gene 5P employed the ellipticity change in the base chromophores at 270 nm induced by binding to fd DNA.[28] As shown in Figure 7, this method can detect binding characterized by dissociation constants less than 10^{-6} *M*. Since the binding constant for the fd DNA complex is $\sim 10^{-9}$ *M*, loss of binding as measured by this assay means a decrease in binding affinity of at least three orders of magnitude. The nitration data suggest that tyrosyl residues 26, 41, and 56 (identified by isolation of the nitrated peptides[28] are in or near the DNA binding site. Susceptibility to nitration does not necessarily indicate a surface position for a tyrosyl residue, nor does failure to nitrate indicate a buried tyrosyl.[68] In this particular case, ^{19}F NMR of gene 5P labeled with *m*-fluorotyrosine does support the division of the tyrosyls into three surface and two buried (see Section III.D).

The susceptibility of gene 5P-nucleotide complexes to dissociation by cations and the distribution of positively charged lysyl and arginyl residues in the N- and C-terminal regions of the molecule has led to suggestions, as with many other DNA-binding proteins, that the positively charged side chains interact with the phosphate backbone of the DNA to form salt bridges in charge-neutralizing interactions. Abolition of binding by acetylation of the lysyl residues supports this postulate.[28] While surface tyrosyls are also *O*-acetylated, the fact that the DNA binding does not prevent the acetylation of the protein strongly suggests that lysyls which remain near the surface in the complex are the primary sites of acetylation. In support of this conclusion, removal of the *O*-acetyl groups from the tyrosyls with hydroxylamine does not restore binding.[28] Since all the lysyl residues are acetylated, no conclusion can be made as to whether specific lysyl residues or all lysyl residues are involved in DNA binding. No chemical modifications have been attempted on the arginyl residues. Proton NMR studies provide some additional information on both lysyl and arginyl interactions in the nucleotide complexes of gene 5P (see Section III.D).

Protection of Cys 33 from mercuration by deoxynucleotide binding does not mean that it interacts directly with the nucleotide. Model building (see below) suggests that

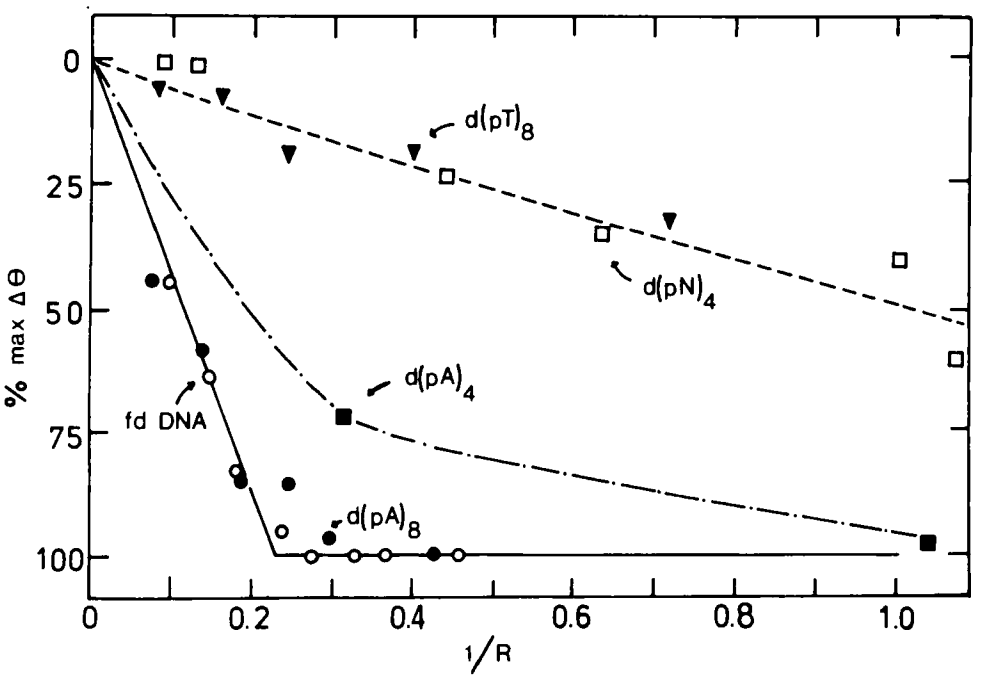


FIGURE 7. Change in nucleotide ellipticity induced by gene 5P (% maximum $\Delta\tau$, at 260 to 270 nm) as a function of the reciprocal of the base: protein ratio, R. O, fd DNA; ●, $d(pA)_8$; ■, $d(pA)_4$; ▼, $d(pT)_8$; □, $d(pN)_4$. Conditions: (1—3) × 10^{-5} *M* nucleotide or DNA (molarity of base); 0.01 *M* Tris-HCl, pH 8.0, 25°C. (Reprinted with permission from Coleman, J. E., Anderson, R. A., Ratcliffe, R. G., and Armitage, I. M., *Biochemistry*, 15, 5419, 1976. Copyright by the American Chemical Society.)

TABLE 2

Summary of Chemical Modification Data

1. Nitration of 3 (Tyr 26, 41 and 56) of the 5 tyrosyl residues of the protein prevents DNA binding; prior binding of DNA or a tetranucleotide completely prevents nitration.
2. Acetylation of the 6 lysyl residues prevents nucleotide binding; acetylation of the lysyl residues in the DNA complex disrupts the complex.
3. DNA binding prevents access of Hg^{++} to the single −SH group (Cys 33).
4. Agents modifying His 64 have no effect on DNA binding.

this residue may be located near the bottom of the DNA binding groove. That this residue is close to the nucleotide bases in the complex has been supported by some UV-induced cross-linking of T-containing deoxynucleotide complexes which show that the nucleotide fragments cross-link to the single cysteinyl residue.[69] Digestion of the cross-linked protein-DNA complex with chymotrypsin, followed by pronase and amino peptidase M yields a nucleotide-cysteine fragment which can be broken by performic acid oxidation to yield cysteic acid.

D. A Multinuclear NMR Approach to the Structure of Gene 5 Protein-Oligodeoxynucleotide Complexes

Most DNA-protein complexes are too large for the application of high resolution NMR techniques. The relatively slow rotational correlation times (near 10^{-7} sec) of molecules of molecular weight greater than 100,000, result in extremely broad signals

which are not resolved by presently available NMR techniques. The small size of gene 5P makes it and its complexes with small oligonucleotides of defined length and sequence an exception. The known amino acid sequence[40-42] and its elaboration in large quantities by a bacterial system also make it particularly accessible to methods required for the isotopic enrichment of specific nuclei of the protein. Thus the gene 5P-nucleotide system provides a prototype to demonstrate the application of methods of high resolution ^{1}H, ^{19}F, and ^{31}P Fourier transform-NMR to the determination of the solution structure of protein-nucleotide complexes.

While the ^{1}H NMR spectrum of gene 5P is relatively highly resolved, the ^{19}F NMR of gene 5P biosynthetically labeled with [^{19}F]-*m*-fluorotyrosine reveals specific information about the environment of the five tyrosyl residues in the protein and their interaction with nucleotides.[31,32] This particular nucleus has the advantage of producing a single well-defined resonance for each tyrosyl residue. The findings complement the tyrosyl chemical modification data discussed above and will be summarized first before the more detailed ^{1}H NMR studies.

1. ^{19}F NMR of Gene 5 Protein Labeled with m-Fluorotyrosine and its Oligonucleotide Complexes

The ^{19}F NMR spectrum of *m*-fluorotyrosyl gene 5P is shown in Figure 8A. There are two resonances, 2.8 and 2.2 ppm, downfield from the chemical shift expected for free *m*-fluorotyrosine, while the other three are grouped between −58 and −59 ppm, approximately the chemical shift expected for free *m*-fluorotyrosine.[70,71] This immediately confirms the conclusion that two tyrosyl residues of the protein are buried (Tyr 34 and 61), their downfield shifts reflecting documented effects on the ^{19}F resonances of fluorotyrosyl residues enclosed within protein structure with alteration of surrounding dielectric constant, more van der Waals contacts, and relative immobilization.[70,71] The other three must be surface tyrosyls and represent the three residues subject to chemical modification (nitration) (Tyr 26, 41, and 56). Resonances of two of the three surface tyrosyls shift upfield when either tetra- or octanucleotides bind, suggesting that it is these three surface tyrosyls which interact with deoxynucleotides. With the binding of $d(pA)_8$, two of the surface tyrosyl ^{19}F resonances shift upfield approximately 0.8 ppm (see assignment in Table 3). A similar pattern of shifts is observed on the binding of $d(pT)_4$, but the maximum upfield shift is approximately 0.4 ppm (Table 3). These upfield shifts are of the right magnitude to represent ring current shifts induced by intercalation of the base rings, and the shifts are greater for adenine than for thymine as expected on theoretical grounds.[72]

The Nuclear Overhauser Effect (NOE) is a modulation of NMR resonance amplitude due to coupling into the spin states of nearby nuclei (in this case protons) which results in an alteration, enhancement, or attenuation, of resonance amplitude when a decoupling frequency is applied. When ^{19}F nuclei are fixed in large slowly rotating macromolecules, the NOE is maximally expressed, and in the case of ^{19}F, ^{1}H is −1. Hence, on proton decoupling of ^{19}F resonances of fixed groups in macromolecules, a frequent observation is complete disappearance of the resonance.[71] As discussed in detail by Hull and Sykes,[71] the expression of the full NOE, and hence the attenuation of the fluorotyrosyl ^{19}F resonances on ^{1}H decoupling, can be modified by internal motion of tyrosyl side chains relative to the protein surface. The NOE of the ^{19}F resonances of fluorotyrosyl gene 5P has been determined, and the expected NOE of resonances 3, 4, and 5 (Figure 8) is not fully expressed (i.e., significant resonance remains on proton decoupling, suggesting that the tyrosyls corresponding to these resonances have rotational motion in addition to that of the gene 5P dimer ($\tau_r = 7 \times 10^{-8}$ sec) and more than the tyrosyl residues represented by the downfield resonances, 1 and 2. This is in accord with the picture of the former three tyrosyls being on the surface of the protein relatively exposed to the solvent.

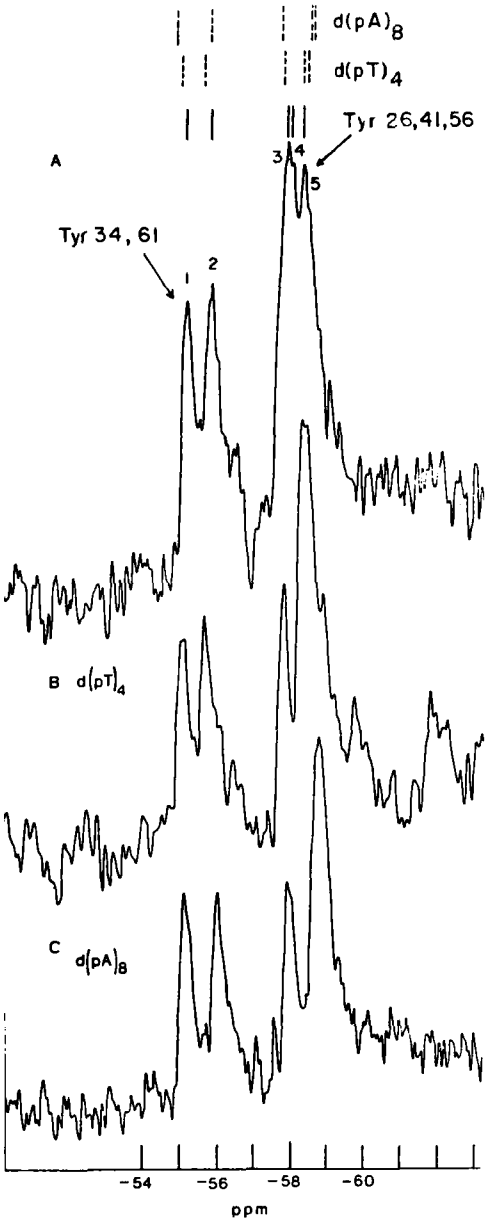
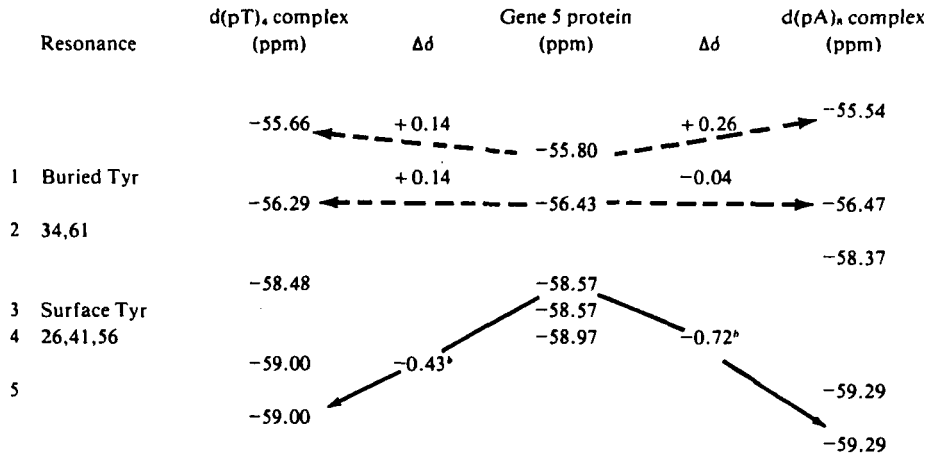


FIGURE 8. ^{19}F NMR of *m*-fluorotyrosyl gene 5P, 5×10^{-4} *M* (A); plus 1×10^{-3} *M* d(pT)$_4$ (B); plus 1×10^{-3} *M* d(pA)$_8$ (C). (Reprinted with permission from Coleman, J. E. and Armitage, I. M., *Biochemistry*, 17, 5038, 1978. Copyright by the American Chemical Society.)

TABLE 3

^{19}F Chemical Shifts[a] in Fluorotyrosyl Gene 5 Protein and its Complexes with d(pA)$_8$ and d(pT)$_8$



[a] Chemical shifts are accurate to ±0.06 ppm. The solid arrows indicate the maximum upfield shifts shown by the fluorines on the surface tyrosyls, while the dotted arrows indicate small downfield shifts that occur on formation of the nucleotide complexes. Conditions: 0.01 *M* Tris, pH 8.0, 10% D_2O, 25°C.

[b] The most upfield ^{19}F resonance position in the complexes is base dependent and contains two resonances (4 and 5) of the same chemical shift. On the other hand, in the unliganded protein, two resonances (3 and 4) have the same chemical shift and are located at the most downfield position of this group. Hence the shifts, Δδ, of −0.43 and −0.72 ppm will apply to at least one and likely both resonances (those labeled 3 and 4 in the unliganded protein). The alternate assignment of resonance 5 in the unliganded protein as one of those moving upfield on complex formation means that it moves upfield 0.3 ppm in the presence of A and only 0.03 ppm in the presence of T, a tenfold difference that appears unlikely. Note that the numbers 1 to 5 refer only to the order of resonances from downfield to upfield.

2. High Field ^{1}H NMR of Gene 5 Protein-Tetra- and Octadeoxynucleotide Complexes

Proton NMR studies of the gene 5P have been carried out at both 270[31,32] and 360 MHz.[73] The proton spectrum of native gene 5P in D_2O at 270 MHz is shown in Figure 9A.[31] The C(2)H resonance of the single histidyl, His 64, is present at 8.02 ppm (peak 1). A number of overlapping aromatic resonances from the five Tyr and three Phe residues and the C(4)H of His 64 appear in the region 6.5 to 7.5 ppm (peaks 3 to 9). The integrated area under these resonances is approximately that expected for 36 aromatic protons, using the single C(2) histidine proton as the standard.

In the highfield region of the spectrum, the ε-CH_2 groups of the 6 Lys residues are resolved in a narrow line at 3.1 ppm. The δ-CH_2 resonances of the 4 Arg residues occur at 3.2 to 3.5 ppm. A large group of overlapping resonances from the methyl groups of the aliphatic residues occurs from 0.5 to 2 ppm, and these are assigned as indicated in Figure 9A. There are a few highfield resonances near 0 ppm (peaks a, b, and c) which represent methyl resonances shifted upfield by ring currents. These assignments are made on the basis of the previous ^{1}H-NMR work on proteins.[74-76]

In the presence of 1 mol of d(pT)$_8$ per 2 mol of protein, significant changes occur in both the linewidth and the chemical shift of the aromatic and aliphatic proton resonances of the protein on formation of the octanucleotide complex (Figure 9B). These changes are best visualized in the difference spectrum A-B. In A-B the resonances

above the base line are those that have disappeared on complex formation, while the inverted cross-lined peaks are new resonances appearing in the complex. The difference spectrum in the aromatic region was determined by assuming that the resonance of the C(2)H of His 64 is unchanged in the complex. In the aliphatic region of the spectrum, the resonance from the lysyl ε-CH_2 groups is unchanged on complex formation and used as a standard. The spectrum of free $d(pT)_8$ under the same conditions, but at three times the concentration, is shown in Figure 9C.

On formation of the octanucleotide complex, peaks 3, 4, and 9 disappear from the aromatic region of the spectrum, while peaks 1, 5, and most of 7 remain unchanged. A part of peak 7,7′, disappears in the spectrum of the complex. Peak 2 in the complex, which contains at least two resonances of different chemical shift, represents a broadened resonance arising from the C(6)H of the thymidine ring.

Since the lines are broadened in the spectrum of the $d(pT)_8$ complex, it is difficult to match exactly resonances disappearing from the spectrum of the free protein with new resonances appearing in the complex. The differential shifts of the aromatic ^1H resonances of gene 5P on complex formation with a pyrimidine- or a purine-containing nucleotide are illustrated by difference spectra for the $d(pT)_8$ and $d(pA)_8$ complex (Figure 10). Comparing these spectra, it is clear that the same set of protons is involved in nucleotide interaction in each case, i.e., those appearing in peaks 3, 4, 7, and 9 in the uncomplexed protein. A group of protons shifted upfield by complex formation (represented by peak 6 in the $d(pT)_8$ complex and peak 11 in the $d(pA)_8$ complex) clearly appears to shift more upfield under the influence of a purine nucleotide than under the influence of a pyrimidine nucleotide (Figure 10).

3. Assignment of the Aromatic Protons in Gene 5 Protein and its Oligonucleotide Complexes

In those proteins where detailed assignments of the aromatic resonances have been made, it has generally been found that the resonances of the 3,5 tyrosyl protons occur to high field (in the range 6.5 to 7 ppm) of the resonances of the 2,6 protons which occur from 7 to 7.5 ppm. In the case of gene 5P, peak 9 must be the 3,5 protons of one of the tyrosyl residues. Tentatively, peaks 7 and 7′ were also assigned to the 3,5 protons of other tyrosyls. Peaks 3 and 4 were originally assigned to the tyrosyl residues and presumably represented the 2,6 protons of the tyrosyl residues interacting with the nucleotide bases.[31]

Growth of an *E. coli* tyrosine auxotroph infected with fd on 2,6-dideuterotyrosine or 3,5-dideuterotyrosine results in the production of gene 5P in which only the 3,5- or the 2,6-tyrosyl proton resonances appear in the NMR spectrum of the aromatic region. The aromatic proton spectrum of gene 5P selectively deuterated in the 2,6-tyrosyl positions is shown in Figure 11 and compared to the spectrum of the fully protonated form. Peak 9 narrows and intensifies, which clearly establishes this resonance as due to 3,5-tyrosyl protons which are now subject to coupling of much smaller magnitude from the adjacent 2,6 deuterons. In addition, reduction of coupling has clearly resolved peak 8 at 6.72 ppm. Since resonance 8 is subject to enhancement by 2,6 deuteration and disappears on complex formation, it can also be assigned to 3,5-tyrosyl protons rather than the histidyl C(4)H. The most striking finding revealed by the spectrum of the 2,6-deuterated derivative is that the 2,6-tyrosyl protons (cross-hatched difference between deuterated and nondeuterated) are all located in the center of the spectrum and contribute only in a minor way to peaks 3 and 4. The latter must be resonances from phenylalanyl protons.

When $d(pT)_8$ is added to the 2,6-dideuterated protein (Figure 11B), resonances represented by peaks 3, 4, 7′, 8, and 9 all move. Peak 6 appears in the complex at the position previously observed in the totally protonated sample (Figure 10). Hence, peak

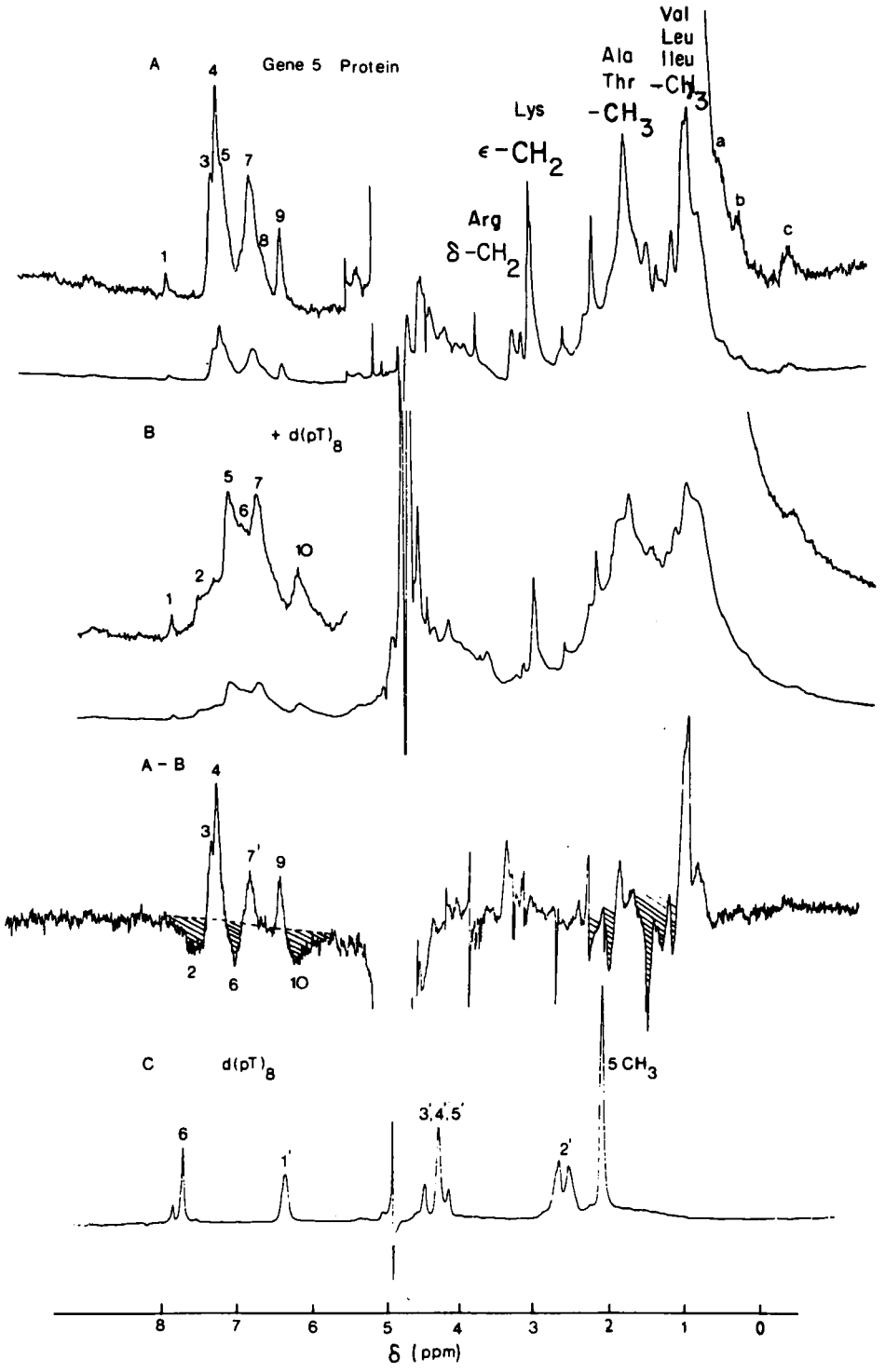


FIGURE 9. 270 MHz ^{1}H NMR of gene 5P, 2×10^{-3} *M* (A); plus 1×10^{-3} *M* d(pT)$_8$ (B). (A-B) is the difference spectrum plotted by setting the C(2)H histidyl resonance (peak 1) and the lysyl ϵ-CH_2 resonance at 3.1 ppm at the same amplitudes in both spectra. (C) 270 MHz ^{1}H NMR of d(pT)$_8$, 3×10^{-3} *M*. Conditions: 0.01 *M* $DPO_4^=$, pH 8.0, 25°C. On the nucleotide spectra, primed and unprimed numbers will refer to the carbon atom carrying nonexchangeable protons in the sugar and base, respectively. (Reprinted with permission from Coleman, J. E., Anderson, R. A., Ratcliffe, R. G., and Armitage, I. M., *Biochemistry*, 15, 5419, 1976. Copyright by the American Chemical Society.)

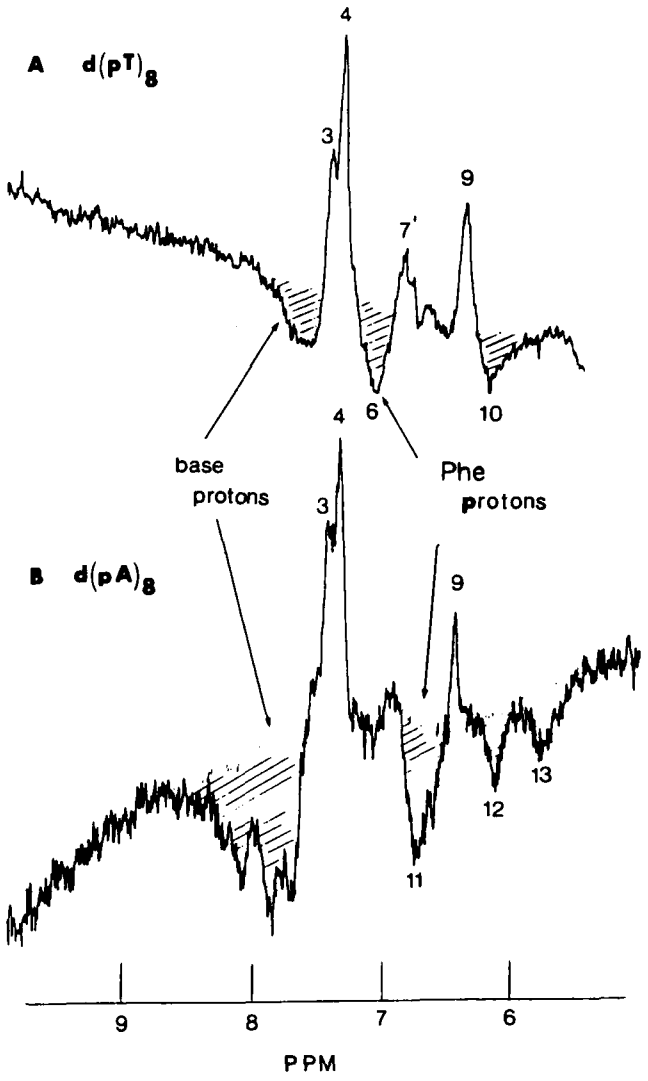


FIGURE 10. ^{1}H NMR difference spectra: A, totally protonated gene 5P d(pT)$_8$ complex; B, totally protonated gene 5P - d(pA)$_8$ complex; □, resonance disappearing from the spectrum on complex formation; ▨, new resonance appearing in the complex. (Reprinted with permission from Coleman, J. E.and Armitage I. M., *Biochemistry*, 17, 5038, 1978. Copyright by the American Chemical Society.)

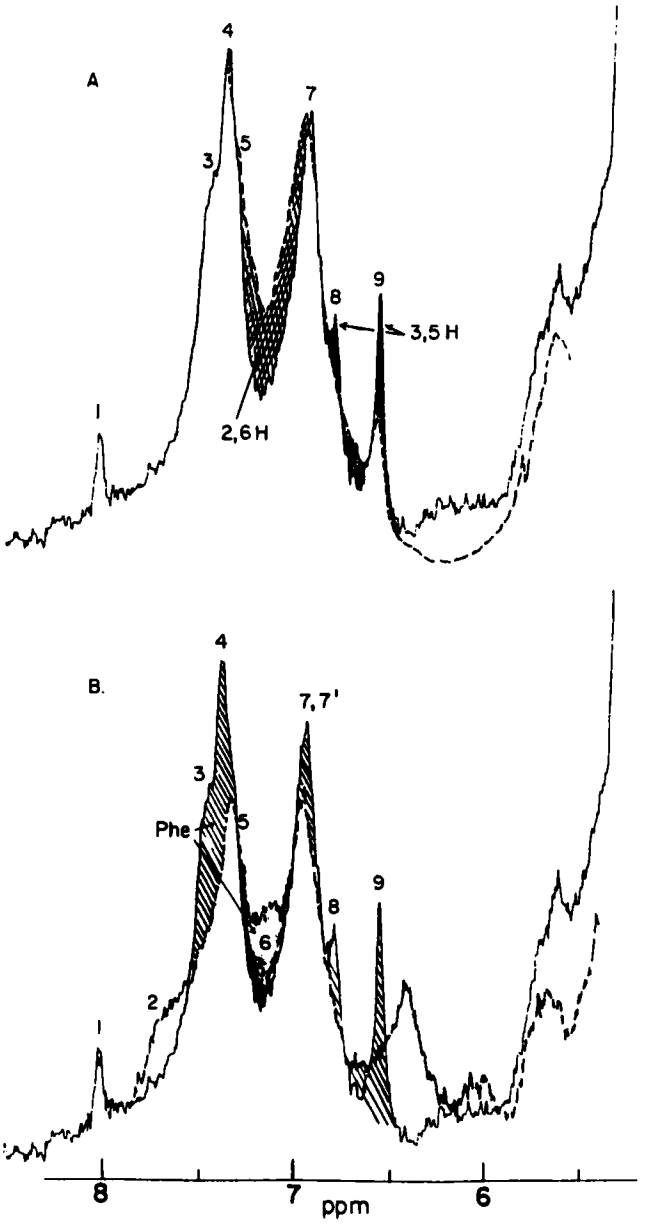


FIGURE 11. (A) (———) Aromatic ^{1}H NMR spectrum of totally protonated gene 5P; (---) spectrum of 2,6-deuterated tyrosyl protein; ■, resonance assigned to 2,6-tyrosyl protons; ■, resonance assigned to 3,5-tyrosyl protons. (B) (—) 2,6-deuterated tyrosyl protein; (----) plus 1 equivalent of $d(pT)_8$; ▨, resonance disappearing from spectrum on complex formation; ■, new resonance appearing in the complex. All spectra have been matched using peak 1, the resonance of the C(2)H of His 64, which is not affected by complex formation. All samples were 10^{-3} *M* protein, pH 8, 20°C. (Reprinted with permission from Coleman, J. E. and Armitage, I. M., *Biochemistry*, 17, 5038, 1978. Copyright by the American Chemical Society.)

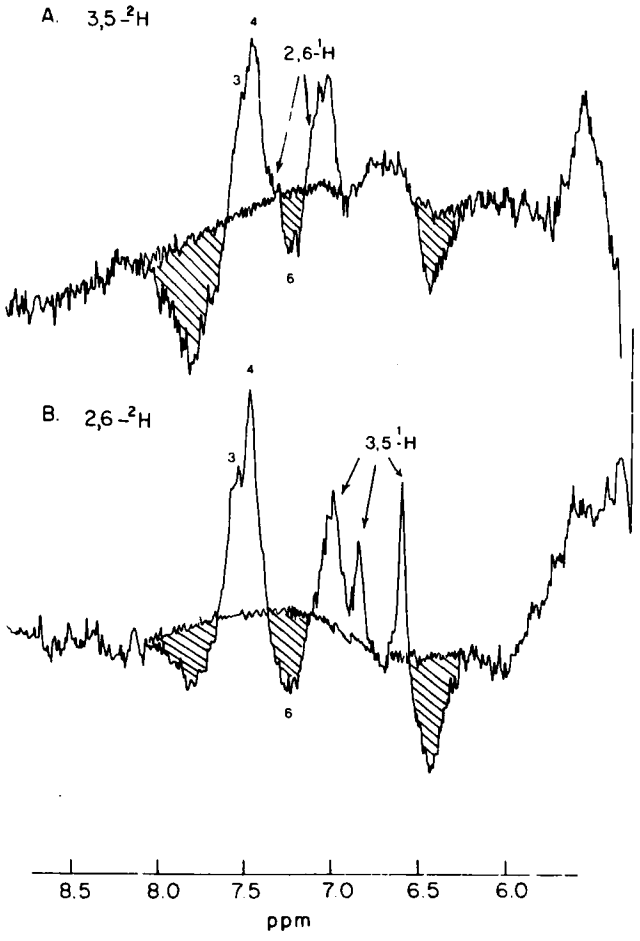


FIGURE 12. ¹H NMR difference spectra: A, 3,5-deuterated tyrosyl gene 5P - 3,5-deuterated tyrosyl gene 5P - d(PT)$_n$ complex; B, 2,6-deuterated tyrosyl gene 5P - 2,6-deuterated tyrosyl gene 5P - d(pT)$_n$ complex; □, resonance disappearing from the spectrum on complex formation; ▨, new resonance appearing in the complex. (Reprinted with permission from Coleman, J. E. and Armitage, I. M., *Biochemistry*, 17, 5038, 1978. Copyright by the American Chemical Society.)

6 in the complex must represent phenylalanyl protons rather than 2,6 tyrosyl protons. Very few, if any, of the 2,6-tyrosyl protons alter their chemical shifts on complex formation. This conclusion is graphically illustrated by the fact that almost all shifts in the proton NMR spectrum observed on complex formation with the fully protonated protein (Figure 10) are also observed on complex formation with the protein in which the 2,6 resonances are absent (Figure 11B).

The quantitative aspects of the qualitative assignment of resonances 3 and 4 to Phe protons can be confirmed by taking the difference spectra between the two deuterated gene 5 proteins and their corresponding d(pT)$_n$ complexes (Figure 12). Both difference spectra show the presence of resonances 3 and 4 in the protein (resonance above the base line) and their replacement by peak 6 in the complex (resonance below the base line). The presence of the 2,6-protons in the spectra derived from the 3,5-deuterated samples (Figure 12A) and the presence of the 3,5-protons in that derived from the 2,6-deuterated samples (Figure 12B) are indicated on the figures. Resonances 3 and 4 belong to neither group, hence they must come from phenylalanyl protons.

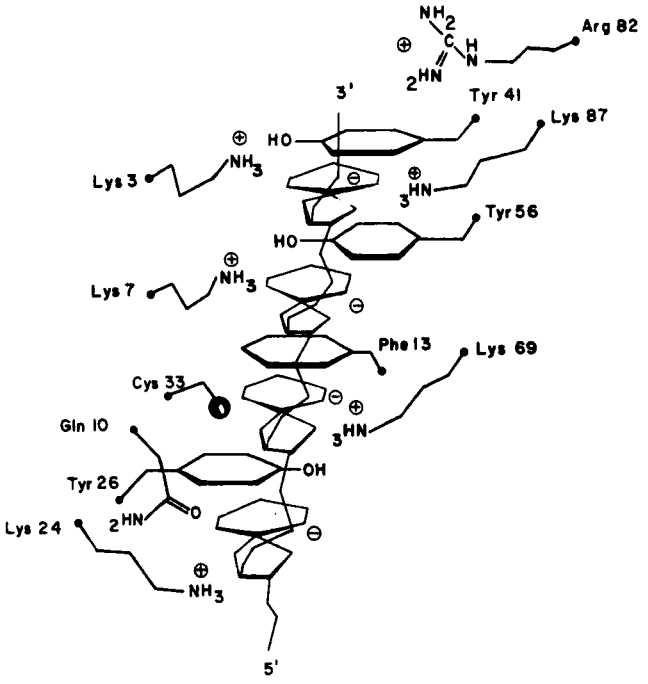


FIGURE 13. Model of the DNA binding surface of gene 5P based on NMR and chemical modification data. (Reprinted with permission from Coleman, J. E. and Armitage, I. M., *Biochemistry*, in press. Copyright by the American Chemical Society.)

On the basis of initial chemical modification and NMR data, a model of the DNA-binding groove was proposed involving intercalation of the bases with tyrosyl residues and electrostatic interactions of the phosphate backbone with positively charged residues of the protein.[31,32] On the basis of the new assignment of peaks 3 and 4 to Phe protons, the model of the DNA-binding surface of the protein has been modified to include both Tyr and Phe residues intercalating with the bases[33] (Figure 13). The three tyrosyls, 26, 41, and 56, are still involved in intercalation, but the rings overlap the base rings only far enough for the 3,5 protons to undergo full ring current shifts. It is possible that one of these tyrosyls does not actually intercalate or even interact directly with the nucleotide, but undergoes changes associated with conformational alteration of the protein on nucleotide binding. This alternative is a possibility, since only two sets of 3,5-tyrosyl protons are clearly involved in shifts (Figure 11), and one of the *m*-fluorotyrosyl resonances of the three surface residues (also sampling the 3,5 positions) moves slightly downfield on complex formation rather than upfield (Figure 8).

From the previous model,[31,32] the Phe 13 is in an ideal position to participate along with the tyrosyls in a base-stacking arrangement. In fact, if Phe 13 is near the ring of Tyr 26, it would explain the abnormal upfield shift of one 3,5 doublet (peak 9, Figure 11) in the unliganded protein. Such an aromatic cluster might also explain the large ellipticity associated with the tyrosyl chromophores.[28] If the bases force these aromatic rings apart, the observed fall in ellipticity on formation of oligonucleotide complexes could be explained. The Phe protons undergo upfield chemical shifts that are greater for purine than pyrimidine bases and are therefore most clearly ring current shifts. The ^{19}F fluorotyrosyl spectrum (Figure 8, Table 3) identifies the 3,5 positions of the tyrosyls as subject to base-dependent upfield shifts of the right magnitude to be ring

current shifts. A ring current shift is one of the few chemical shifts that is independent of nucleus.

Completing the multinuclear NMR approach to the structure of gene 5 protein oligonucleotide complexes, the ^{31}P NMR has been used to investigate the conformation of the (phosphodiester) backbone of the bound nucleotide. Upon binding, the 3′-5′ diester resonance moves upfield in both $d(pT)_8$ and $d(pA)_8$ complexes, suggesting that the diester conformation in the complex may be more like that found in helical conformations of DNA, the gauche-gauche rather than the trans-gauche preferred in the random conformation. The maintenance by HDPs of a phosphodiester backbone conformation of the single strand, which is similar to that found in helical polynucleotides, could have significant functional implications. Subsequent transformation of these complexes either by enzymes, e.g., polymerases or other packaging proteins, might be facilitated by the presence of the helical conformation in the single strand.

E. Crystal Structure of Gene 5 Protein

Crystals of gene 5P suitable for X-ray diffraction studies with a resolution of 2.3 A have recently been obtained by the vapor diffusion technique using solutions of polyethyleneglycol.[43,77] The fundamental unit of quaternary structure observed in the electron density map determined by McPherson et al.[43] is a dimer with molecular twofold symmetry. The monomers are held together by the interlocking of complementary regions of β-loop which are located a considerable distance from that part of the molecule believed to serve as the DNA-binding groove. It is reasonable to conclude that this dimer represents the physiologically important unit of gene 5P. The binding of two oppositely directed strands of fd DNA to each of the constituent monomers would explain the collapsed DNA-complex structure observed in the electron micrographs (Figure 1). A tentative model of the tracing of the polypeptide chain is shown in Figure 14B. This is derived from the best fit of the electron density map at the current stage of refinement, but the details could be modified with further refinement. The dimensions of the monomer are approximately 45 Å × 25 Å × 30 Å.

The secondary structure of the protein appears to be entirely antiparallel β-structure interposed with short lengths of extended chain. There appears to be no α-helix. There is thus general agreement with the predictions from sequence-structure rules as outlined above. The arrangements of the β-pleated sheet into antiparallel strands is diagrammed in Figure 14A. Beginning about ten residues from the N-terminus, there is a three-stranded antiparallel β-sheet. The chain then loops out to form the short two-stranded antiparallel β-sheet located astride the diad axis which constitutes the main feature of the monomer-monomer interlock. The remaining β-structure forms a two-stranded antiparallel sheet lying on top of, but at an angle to, the three-stranded sheet. For a distance of ten or so residues, both the N- and C-terminal portions of the polypeptide chain appear to be wrapped on the molecule in relatively extended form.

The three-stranded β-sheet is not a flat surface, but like many short β-sheets is radically twisted, which produces a concavity or corridor along the underside of the molecule. The tunnel-like nature of this cavity is enhanced by the placement of the adjacent two-stranded β-sheet which juts at an angle of 120° to the three-stranded sheet and forms the interaction surface with the adjacent monomer. The long groove beneath the three-stranded β-sheet, by its shape and extent (∼30 Å), would suggest that this structure forms the DNA-binding region. Comparing the amino acid residues that make up this region to those identified by chemical modification and NMR methods to interact with nucleotides (Sections III.C and D), one finds that Tyr 41 defines one end of this concavity, while Tyr 26 defines the other. In addition, Phe 13 occupies this region and Cys 33 lies behind this groove. Tyr 56 also lies in this region as do two

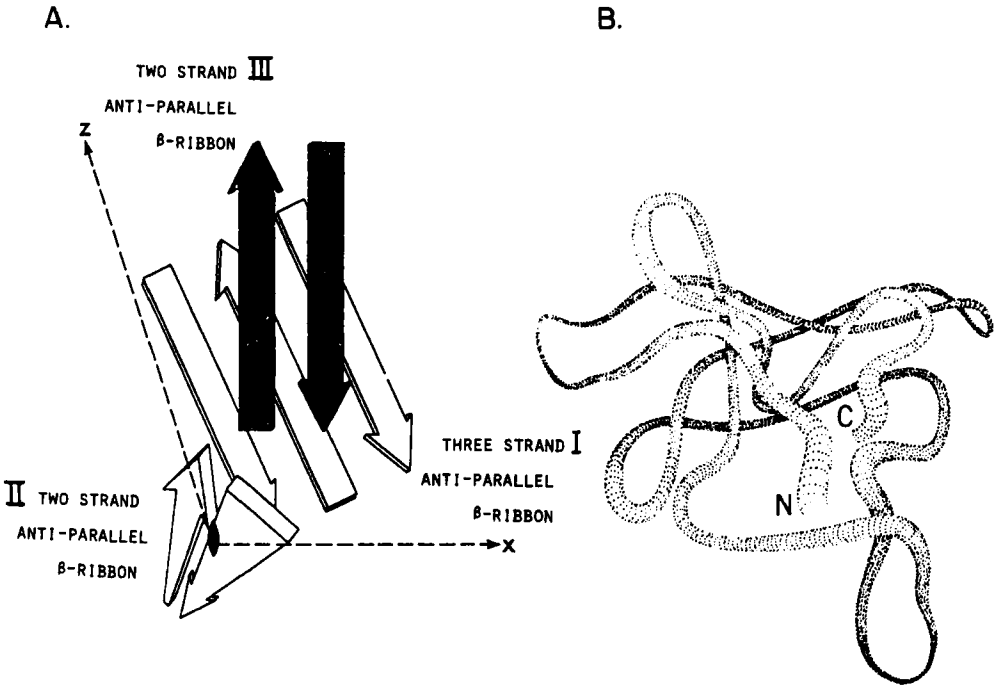


FIGURE 14. (A) A schematic diagram showing the development of the secondary structural components of the gene 5P molecule. The three strands of the antiparallel β sheet (I) form the major part of the DNA binding interface. The two strands of the β ribbon (II) appear to be principally responsible for maintaining the molecule as a dimer in solution by its interaction with a symmetry-related β ribbon. The second β ribbon (III) is established diagonally across the β sheet (I). β ribbon III may be the primary participant in the lateral interactions from which the cooperativity of DNA binding arises. (B) A schematic three-dimensional drawing showing the course of the gene 5P polypeptide backbone taken directly from the electron density map. The molecule appears to be composed entirely of β structure and extended chain with no α helix. (From McPherson, A., Jurnak, F., Wang, A., Kolpak, F., Rich, A., and Molineux, I., *Cold Spring Harbor Symp. Quant. Biol.*, in press. With permission.)

arginyl residues, but precise relationships to a bound nucleotide cannot be determined in the absence of a crystal structure for a nucleotide complex.

The findings on the unliganded molecule have been summarized by McPherson et al.[43] in terms of three functional domains. The three-stranded antiparallel β-sheet (I) is primarily responsible for binding to and interacting with the DNA. The first two-stranded antiparallel β-ribbon (II) maintains the dimer in solution by tightly interlocking with a symmetry-related loop on the other monomer. The second β-ribbon (III) is postulated to be involved in cooperative side-to-side protein interactions to neighboring gene 5P dimers, although part of this structure overhangs the three-stranded DNA-binding region, and the tip, or bend, could contribute residues to the DNA-binding interface. Thus, in a very economical fashion, gene 5P accomplishes three major functions: DNA binding, dimerization, and side-to-side oligomerization, with the folding of a polypeptide chain of only 87 amino acid residues.

1. Crystals of the Gene 5 Protein-Oligonucleotide Complexes

It has not been possible to diffuse small oligonucleotides into the presently available crystal form of the unliganded gene 5P and have them bind. This is probably due to the molecular packing arrangement which stacks the molecules closely along the y axis of the crystal with a 2, screw symmetry. The result of this tight packing is to fill and

block the presumptive DNA-binding groove with density from neighboring molecules. Hence, the precise interactions responsible for nucleotide binding have not been determined.

While oligonucleotides do not appear to bind to gene 5P in the crystal of the initially unliganded molecule, crystallization by the vapor diffusion technique from polyethylene glycol of solutions of gene 5P containing a variety of oligodeoxynucleotides of defined length and sequence results in crystals of different morphology and molecular packing from that observed for gene 5P alone (termed the native crystal). There seems to be little dependence of crystallization on the length or specific sequence of the oligonucleotide present. Similar diamond-shaped plates of complex have been grown from gene 5P solutions containing d(GpC), d(ApT), $d(Ap)_4$, $d(Ap)_8$, and d(CpTpTpCp). Hexagonal prisms crystallize from solutions of gene 5P and hexanucleotides.

While detailed work on any of these crystals has only just begun, it is probable that these are crystals of the complexes. This is also supported by the finding that only native crystals form if the oligonucleotide to protein ratio is less than 1.0, while above that ratio, only the complex crystals form.

Preliminary diffraction data have been collected on several of the complex crystals, and all demonstrate hexagonal symmetry, and the asymmetric unit weight calculates to be about 12 gene 5P monomers or 6 dimers. This hexameric array of dimers is postulated to be the structure induced by the binding of oligonucleotides and may bear a relationship to the 12 gene 5P monomers (or 6 dimers) per turn of the helix believed to occur in the gene 5P-fd DNA complex based on measurements from electron microscopy.[62] The occurrence of six dimers per asymmetric unit suggests the presence of a sixfold axis. Since the crystal contains fixed aggregates and not infinite strands, the arrangement of dimers must be closed, and the likely arrangement is a circle of dimers with a natural sixfold axis along its center, perpendicular to the twofold axes of the dimer units, i.e., 622 point group symmetry. Such a structure of gene 5P dimers is pictured in Figure 15A and is postulated to be the arrangement induced by the binding of nucleic acid fragments. The aggregate would have a twofold crown shape of diameter $\sim$100 Å and a thickness of $\sim$80 Å. One model of how such a structure might be altered by complex formation with two antiparallel DNA chains producing infinite side-to-side interactions is pictured in Figure 15B. This "lock washer" structure results from opening the ring of dimers and displacing the ends along the direction of the sixfold axis. An infinite stack of these would produce the desired superhelix, maintaining essentially the same contacts between adjacent monomers as observed in the closed structure in the crystal. McPherson et al.[43] point out that the gene 5P aggregate could, in principle, form with the DNA binding either on the outside or inside of the ring. Based on the analysis of the tertiary structure of the monomer, however, the molecular relationships appear to fall into place much better if the dimers are placed such that the DNA binding grooves are on the outside. The two DNA single strands are then wrapped around the gene 5P core at a radius of 30 to 40 Å, a reasonable radius for DNA coiling. The alternative model would require unreasonably tight DNA coiling.

To quote McPherson et al.[43]

> a mechanistic advantage of this model is that it provides a simple means for the displacement of the gene 5P by the gene 8 coat protein. Since the DNA strands are on the exterior of the gene 5 complex, they are exposed to the approach of the coat protein which also binds to DNA cooperatively, but much more tenaciously than the gene 5 protein. When the gene 5/DNA complex reaches the cell membrane, the coat protein binds very strongly to the DNA strands on the outside of the complex. Since the gene 8 protein is of a much smaller size, it undoubtedly forms a helix with different parameters. The binding of the coat protein would probably induce a conformational change in the DNA and thereby disrupt the gene 5/DNA bonds as well as the gene 5P dimer-dimer interactions. Thus, the gene 5 protein which forms a spindle on the interior of

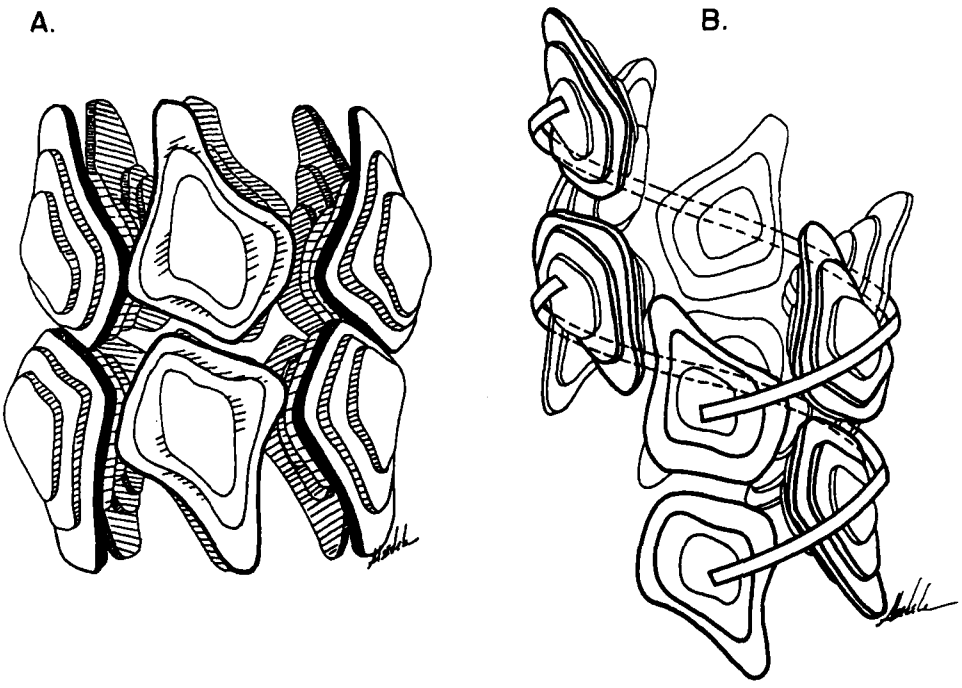


FIGURE 15. (A) A proposed model for the asymmetric unit common to the four crystal forms of gene 5P-DNA complexes. The arrangement is a circle, or disk, having 622 point group symmetry formed by joining the two ends of a linear array of six gene 5P dimers. The upper hexagon of monomers will bind a single strand of DNA running in one direction and the bottom level, a strand running in the opposite direction. The disk has a diameter of about 100 Å and a height of about 80 Å. (B) A proposed model for the structure of one turn of the gene 5P-DNA 2-stranded helical complex. This arrangement arises by opening the disk structure seen in part A between any two adjacent dimers and displacing the free ends along the unique axis direction. The stacking of these "lock washer" units results in a double helical structure having a sixfold screw axis with perpendicular dyads, twelve gene 5P monomers per turn, and dimensions consistent with the helices observed by electron microscopy. The two DNA single strands are spooled around this spindle of gene 5P. (From McPherson, A., Jurnak, F., Wang, A., Kolpak, F., Rich, A., and Molineux, I., *Cold Spring Harbor Symp. Quant. Biol.*, in press. With permission.)

the cell is exchanged for a sheath of coat protein which must protect the virion after extrusion into the media.

IV. *E. COLI* HELIX-DESTABILIZING PROTEIN

A helix-destabilizing protein (originally called *E. coli* "unwinding" protein) with properties rather similar to the T4 gene 32P can be isolated from *E. coli*. Using single-stranded DNA cellulose chromatography, the protein can be purified to homogeneity.[21,78,79] Its molecular weight on denaturing gels has been reported as 22,000 with a native molecular weight of 90,000 suggesting a tetrameric structure.[79] No higher aggregates were found. Weiner et al.[80] have taken advantage of the property of this protein to stimulate the in vitro conversion of single-strand G4 phage DNA to the replicative form to devise simple assay for the protein and have purified the *E. coli* HDP to homogeneity by a different procedure. The method includes a 100°C heating step to which the protein is resistant. Their preparation shows a single band of 18,500 daltons on denaturing gels and a native, tetrameric, molecular weight of 76,000 ± 4000. The origin of the discrepancy in molecular weight is unclear. Weiner

et al.[80] have characterized the protein extensively. It is an asymmetric tetramer with an isoelectric point of 6.0.

The protein binds tightly to single-stranded DNA with a stoichiometry of one tetramer per ~30 bases as judged by various assays.[79,80] Complexes of *E. coli* HDP with single-stranded phage DNA and partially denatured duplex DNA can be visualized by electron microscopy[78] (Figure 1). Like the complex formed by gene 32P, when fd DNA is used, the circle remains open upon complex formation with *E. coli* HDP (Figure 1). The nonrandom distribution of protein on the fd DNA circle at less than saturating concentrations indicates highly cooperative binding, as in the case of T4 gene 32P.[19] As would be predicted from its properties, this *E. coli* protein promotes the denaturation of duplex T4 DNA. This can be readily followed in the spectrophotometer, since the DNA-protein complex is fully hyperchromic. Catalysis of DNA renaturation by the *E. coli* HDP is, however, a more complex matter.[81] It occurs at low pH in the presence of divalent cations, but at pH 7, the catalysis of renaturation appears to be dependent on polyamines. It remains possible, therefore, that the *E. coli* HDP has a role in recombination, as does gene 32P. There are estimated to be approximately 800 copies of this protein in a rapidly dividing *E. coli* cell.[78]

The *E. coli* HDP is capable of stimulating homologous DNA polymerases. Sigal et al.[78] demonstrated a stimulation of DNA polymerase II on extensively gapped templates. Since then, this HDP has been shown to be required for maximum activity in a variety of in vitro viral DNA replication systems, both in viral strand to replicative form conversion[60,79,80,82,84,85] and in replication of viral duplex DNA.[87] These findings imply a stimulation of DNA polymerase III, the enzyme used in *E. coli* DNA replication.

It is probable that the function of *E. coli* HDP in the presynthetic stage of the conversion of the infecting single strands to the replicative form is to mask the single-stranded DNA and isolate a small duplex hairpin at which priming of DNA synthesis by an RNA polymerase priming enzyme can exclusively occur.[60,82,83,86] *E. coli* does indeed inhibit RNA polymerase activity on single-stranded templates.[79] In the elongation reaction, the HDP presumably acts in a manner analogous to that proposed for gene 32P, binding to the template and holding it in a conformation favorable to the DNA polymerase. In addition, it is suggested[87] that during rolling circle replication it may bind to progeny single strands. A diagram of the complete in vitro system that carries out the efficient conversion of the single-stranded viral strands of fd (or M13) DNA to the double-stranded replicative forms is given in Figure 16, based on the findings of Kornberg and collaborators (for detailed discussion see Kornberg[45] and Lewin[88]).

These findings suggest that the *E. coli* HDP is required for replication of the *E. coli* chromosome, and a rough calculation from the amount in a cell and the average number of replication forks suggests that approximately the same number of HDP molecules per fork would be present as with the gene 32P in T4 replication.[78] Mutations of the gene for the *E. coli* HDP have not been detected, suggesting that most mutations altering the function of the protein may be lethal.

The yields of *E. coli* HDP are much less than those realized from preparations of the phage specific HDPs, hence much less investigation of the physicochemical properties of this protein has been carried out. The circular dichroism of the peptide bond chromophores shows a mean residue ellipticity remarkably similar to that of gene 32P.[21] Analysis indicates a similar mix of α, β, and random structure.[21] There is in addition some negative ellipticity centered at 280 nm associated with the aromatic chromophores.

The ellipticity changes induced in the base chromophores of fd DNA by the *E. coli*

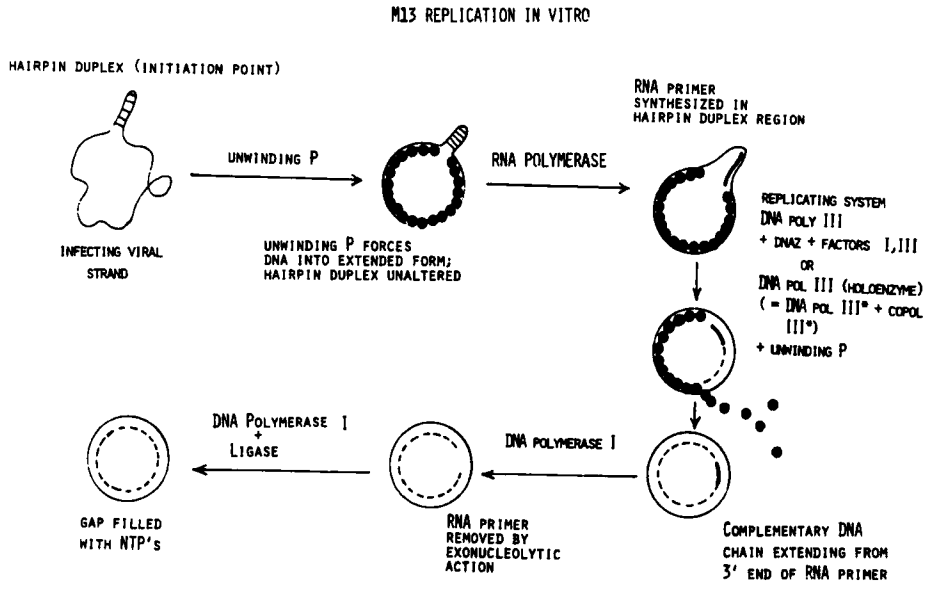


FIGURE 16. Schematic diagram of the role of *E. coli* HDP in the in vitro conversion of single viral strand of fd(M13) to replicative form.

HDP are very similar to those produced by gene 32P, a fall and reversal of sign of the bands in the 260- to 280-nm region and a movement of the cross-over from 260 to ~290 nm.[21] The circular dichroism change is complete at a nucleotide to protein monomer ratio of ~14:1,[21] which is significantly different from eight bases per monomer estimated from the hyperchromism induced in DNA by the protein[78] or as calculated from sucrose density centrifugation of the complex.[78] A nucleotide to tetramer ratio of 30—36 has been determined by gel filtration of the complex formed between ϕX DNA and ^{125}I-labeled *E. coli* HDP.[80]

V. T7 HELIX-DESTABILIZING PROTEIN

With the discovery that the T4 gene 32P and the *E. coli* HDP bound to single-stranded DNA and stimulated their homologous DNA polymerases came the suggestion that each DNA polymerase required its own conjugate HDP for full activity. The HDP presumably would coat the single-strand template, removing obstructing intrastrand hairpins, and holding the template in a favorable conformation. The phage T7 codes for its own DNA polymerase for replication[89] and this led to independent searches and subsequent isolations of a T7 HDP analogous to gene 32 protein.[90-92]

Using chromatography on single-stranded T7 DNA-cellulose, a T7-coded HDP can be purified from T7-infected *E. coli* between 5 and 18 min after infection. Present in large amounts (0.1% of all soluble protein), this protein has a molecular weight of between 25,000[92] and 31,000.[90,91] In sedimentation assays, the protein indeed binds exclusively to denatured DNA, and it protects duplex DNA from exonuclease III digestion at 37°C, suggesting that melting of the DNA is accomplished at this temperature.[92]

The T7 HDP stimulates T7DNA polymerase activity on single-stranded templates, especially at low temperatures where in its absence, activity is low, presumably due to obstructing intrastrand hairpins in the template.[91] Stimulation is also observed on "gapped" templates.[90-92] *E. coli* HDP can substitute for the T7 protein with T7 DNA polymerase in vitro,[91,92] but T7 HDP cannot substitute for the *E. coli* protein with the

host DNA polymerase II.[90] Extensive investigations of the T7-specific HDP have not been carried out as yet.

VI. EUKARYOTIC HELIX-DESTABILIZING PROTEINS

HDPs have now been identified in eukaryotic cells varying from plants and fungi to man. We will review these proteins and will stress the extent of their analogy to prokaryotic HDPs, in particular to gene 32P as far as this has been investigated in each case. Characteristics of interest will be (a) preferential binding to single-stranded DNA, (b) lowering of T_m of duplex DNA, (c) holding SS-DNA in an extended conformation and promoting renaturation under the correct conditions, and (d) stimulation of the homologous DNA polymerase on appropriate templates.

A. Helix-Destabilizing Protein from Meiotic Cells

Hotta and Stern[93,94] isolated a protein similar in many respects to gene 32P from microsporocytes of lily during meiotic prophase. The protein can be partially purified from nuclei by sucrose gradient centrifugation and has a density characteristic of a lipoprotein. The lipid can be removed with detergent without affecting its DNA-binding properties as measured with a nitrocellulose filter binding assay in which, in contrast to the usual finding, the protein-DNA complex passes through the filter and the unbound single-strand DNA remains bound.

After further purification on DEAE cellulose, the molecular weight of the protein was estimated at 35,000, very similar to that of gene 32P. Its binding is specific for single-stranded DNA, with an absolute requirement for a divalent cation. Also in common with gene 32P, the lily protein catalyzes the renaturation of thermally denatured lily and T7 DNA.

The lily protein is present only in nuclei in meiosis, appearing transiently during prophase. It is absent from cells in other phases of meiosis and from all somatic tissues. The time of its appearance and its similarities to gene 32P suggest a role in chromosome pairing and crossing over in zygotene and pachytene.

Nuclei in meiotic prophase from rat, bull, and human testes yield similar DNA-binding proteins.[95] Isolated as lipoproteins, these species bind specifically to single-stranded DNA and catalyze the renaturation of thermally denatured DNA. This component is absent from mitotic testicular cells, liver cells, fibroblasts, and lymphocytes.

With the advent of single-stranded DNA-cellulose affinity chromatography, Mather and Hotta[96] were able to isolate the meiotic DNA-binding protein of rat spermatocytes (known as an R protein to emphasize its proposed role in recombination) in a pure, lipid-free state. Its molecular weight is 33,000 to 38,000. In the absence of Mg^{2+}, the R protein promotes the unwinding of duplex DNA and the renaturation of heat-denatured DNA. These activities could be inhibited by phosphorylation of the protein with protein kinase. If serine residues are the sites of phosphorylation, it can be calculated that phosphorylation of 10 to 12 residues per molecule abolishes activity. Subsequent treatment with alkaline phosphatase restored both unwinding and renaturation activities. The protein as isolated is phosphorylated to some extent, and its activity increased on dephosphorylation.[96] These studies suggest that in vivo phosphorylation of DNA-binding proteins may play an important regulatory role in meiosis.

B. Helix-Destabilizing Proteins from Cultured Mammalian Cells

Early studies on the DNA-binding proteins of cultured cells were directed toward variations of these proteins during the cell cycle and their possible role in controlling DNA synthesis. Using single-stranded DNA-cellulose chromatography, Salas and

Green[97] identified eight DNA-binding proteins, P1 to P8, in 3T6 mouse fibroblasts. One of these, P6, is present predominantly in exponentially growing cells, in S phase, and therefore may be involved in DNA synthesis. Another protein, P8, increases when DNA synthesis is arrested by a thymidine block. Two of the others were subsequently shown to be collagen precursors having a high affinity for DNA.[98] Rubio et al.[99] and Rubio and Long [100] have studied the changes in these DNA-binding proteins when cells are transformed by tumor viruses. 3T6-MSV(RLV) produced increased amounts of P3, P5, and P6. 3T3 cells produced P8 at the time of initiation of DNA synthesis, and MSV-transformed 3T3 cells in serum-depleted media still produced P8 at this time. This evidence strengthens the hypothesis that P8 functions in DNA synthesis.

Tsai and Green[101] have purified P8 from 3T3 cells using DNA cellulose chromatography. The molecular weight of P8 as estimated on denaturing gels is ~20,000. In contrast, gel filtration suggests a much higher molecular weight. P8 would therefore appear to form oligomeric structures like gene 32P and the *E. coli* HDP, which show self-association under certain conditions (see above). Of all the DNA-binding proteins of the cell, only P8 is absolutely specific for single-stranded DNA. In contrast to gene 32P, P8 does not promote renaturation of denatured DNA, and its role in the cell may not be exactly analogous to that of the phage protein.

DNA-binding proteins of Chinese hamster ovary cells were studied in a similar manner by Fox and Pardee.[102] No differences were detected in these proteins through the cell cycle, although in serum-depleted cells the DNA-binding proteins differ. The DNA-binding proteins of the human fibroblast line, W138, have been studied.[103,104] Stein[105] has characterized one of them, a 33,000-dalton protein, P8, and it seems to be comparable to the mouse fibroblast, P8. It, too, binds preferentially to single-stranded DNA.

C. Helix-Destabilizing Proteins from Virus-Infected Cells

Van der Vliet and Levine[106] and Shanmugam et al.[107] have used single-stranded DNA-cellulose chromatography to isolate HDPs from cells infected with adenovirus. Two such proteins of 72,000 and 48,000 daltons (75,000 and 45,000, in the estimate of Shanmugam et al.) were found that have properties in common with gene 32P. They bind only to single-stranded DNA (although not RNA) and will accept not only adenovirus DNA, but also fd and calf thymus single-stranded DNA. Cells infected with a virus with a temperature-sensitive mutation in DNA synthesis lack these DNA-binding activities, suggesting a role for them in DNA synthesis.

In cells infected with polyoma virus, Yeh et al.[108] have detected a protein that nonspecifically binds duplex DNA, as judged by a nitrocellulose filter assay. Its binding to single-stranded DNA was not tested. The protein does, however, have one property in common with gene 32P in that it stimulates DNA polymerase activity. The polyoma-induced protein stimulates *E. coli* (Pol I) and T4 DNA polymerase activity on single-stranded and double-stranded templates.

D. Helix-Destabilizing Proteins from Primitive Eukaryotes

Helix-destabilizing proteins have been identified to date in two primitive eukaryotes, *Ustilago maydis* and *Tetrahymena pyriformis*. Banks and Spanos,[109] using single-stranded DNA-cellulose chromatography, purified a HDP from *U. maydis*. It could be purified to homogeneity on DNA-cellulose, and had a molecular weight of 20,000 on denaturing gels. Molecular weight estimates from glycerol gradient sedimentation suggest that no self-association occurs in the case of the *U. maydis* protein. DNA binding is specific for single-stranded DNA, as assayed with a nitrocellulose filter binding assay. The sigmoid binding curve perhaps indicates a cooperative binding mode.

The binding site for the protein appears to be seven to ten single-stranded nucleotides by calculation from the binding curve. This protein promotes the denaturation duplex DNA (reducing the melting temperature of poly [d(A-T)] · [d(A-T)] by 50°C) and the renaturation of thermally denatured DNA at physiological temperatures.

In striking comparison to the prokaryotic HDPs, the *U. maydis* protein stimulates the activity of *U. maydis* DNA polymerase on partially single-stranded templates. No heterologous polymerase was stimulated. No association between the HDP and the homologous polymerase was detectable, however, and it seems to function primarily by binding to the DNA and perhaps by holding the template in a favorable conformation. The presence of this protein in large amounts (about 3×10^5 per cell) in mitotic cells of *Ustilago* is in contrast to the meiotic R-proteins described above, and suggests a different role in vivo for this protein.

Three proteins (A, B, and C) that bind specifically to single-stranded DNA have been isolated from *T. pyriformis* by Donnell et al.,[110] using single-stranded DNA-cellulose chromatography. The molecular weights of these proteins are 47,000, 41,000, and 32,000, respectively. The A protein was studied in detail and has much in common with the gene 32P. It binds preferentially to single-stranded DNA in a cooperative manner. At high concentrations of protein and low ionic strength, there appears to be an interaction between A protein molecules that precludes binding to oligo $(dT)_{100}$.

E. Helix-Destabilizing Proteins from Normal Mammalian Tissue

1. Calf Thymus

The most extensively investigated mammalian tissue is calf thymus. Herrick and Alberts[111] have isolated from homogenates three proteins with affinity for single-stranded DNA and not duplex DNA. Their fractionation procedure, which is of general use, selects HDPs by passing them through a native DNA-cellulose column, binding them to a single-stranded DNA column, and competing out proteins with a nonspecific affinity for polyanions with sodium dextran sulfate. Proteins remaining on the column can be eluted with progressively increasing salt concentration. Three proteins can be purified from calf thymus in this manner. The predominant species is a 24,000-dalton monomeric HDP, designated calf thymus unwinding protein 1 (UP1). This, however, is resolved by isoelectric focusing into four or five subspecies which have different DNA affinities. The remaining two proteins both have molecular weights of 33,000 in denaturing gels. One, a basic protein, elutes at low salt and the other, an acidic protein, at high salt.

Both UP1 and the high-salt-eluting species promote the melting of poly[d(A-T)] and poly[r (AU)] (the latter indicating an affinity for RNA) and of natural DNA.[112] They also destabilize hairpin helices in single-stranded DNA, although they do not promote renaturation of denatured DNA, a significant difference between them and gene 32P. It is unclear whether or not binding is cooperative.[112]

It is generally assumed that the helix-destabilizing proteins melt DNA simply because of their affinity for single-stranded DNA and the subsequent displacement of the helix-coil transition. In the case of UP1, however, a significant affinity for duplex DNA has been demonstrated by a competition experiment,[112] suggesting the possibility of an active protein-mediated denaturation.

In the matter of stimulation of the homologous DNA polymerase, UP1 is analogous to T4 gene 32P. On a partially single-stranded calf thymus DNA template, the activity of calf thymus DNA polymerase is stimulated tenfold under conditions where the single-stranded regions are saturated with UP1.[113] No stimulation is observed with gene 32P, suggesting again that the HDPs hold the template in a conformation favorable specifically to the homologous polymerase.

2. Rat Liver

Two studies have identified HDPs in rat liver. Thomas and Patel[114] further fractionated a fraction of nonhistone chromosomal protein with single-stranded DNA-agarose chromatography and discovered a 20,000-dalton protein which exhibits preferential binding to single-strand DNA in a nonspecies-specific manner and depressed the melting temperature of poly[d(A-T)] · [d(A-T)], such that melting occurred at 25°C.

In regenerating rat liver, a different HDP can be demonstrated in the cytosol.[115] Present in approximately 10^6 copies per cell, the protein is a tetramer with subunits of 25,000 daltons. Again, a depression of the T_M of poly[d(A-T)] was observed due to preferential binding to single-stranded DNA, and in common with many others discussed, this protein stimulates homologous DNA polymerases α and β on partially single-stranded templates.

F. Helix-Destabilizing Proteins from Transformed Cells

An interesting HDP, probably corresponding to the acidic 33,000-dalton protein from calf thymus,[111] has been isolated from mouse ascites cells.[116] The purified protein has a molecular weight of 30,000 to 35,000 and binds preferentially to single-stranded DNA, maintaining it in an extended conformation. It stimulates the homologous DNA polymerase-α on single-stranded templates. The protein can be modified by phosphorylation with protein kinase at probably a single site. This modification does not affect the DNA-binding affinity, but does abolish the effect on DNA polymerase.

Mouse cells transformed by chemical carcinogens are also a source of a single-stranded DNA-binding protein.[117] The protein has a molecular weight of 25,000 to 30,000, but was not characterized further.

Finally, human chronic lymphocytic leukemia lymphocytes have been shown to contain a HDP of 25,000 daltons (probably a dimer of subunits of ~13,000 and ~11,000 daltons).[118] It binds tightly to single-stranded DNA and to UV-irradiated DNA, promoting both helix-coil transition and incision by UV endonuclease. This latter finding suggests a cause for the increased resistance of these neoplastic cells to radiation and emphasizes the similarity of this protein to gene 32P, which has been shown to be required for UV repair in T4 in vivo.[12]

VII. IMPLICATIONS OF BINDING MECHANISMS OF HELIX-DESTABILIZING PROTEINS FOR THOSE OF PROTEINS RECOGNIZING SPECIFIC DNA SEQUENCES

It is likely that the study of the helix-destabilizing proteins will yield general principles and mechanisms applicable to other proteins that bind to DNA. Proteins currently under intensive investigation would include the histones,[119] nonhistone chromosomal proteins,[120] DNA polymerases,[45] topoisomerases,[121-123] and restriction endonucleases.[124]

Among the proteins crucial to gene expression and its control are those that bind to specific DNA sequences occurring once or a few times per genome. Examples of such proteins are RNA polymerases, repressors, and CAP protein. If we accept the proposition that such binding has two components, one a general DNA-binding function and the other conferring sequence-specificity, then lessons learned in studying the helix-destabilizers might be applicable to the former aspect of these proteins as well. To demonstrate this, we will briefly discuss three such proteins, the RNA polymerases of *E. coli* and T7, and *lac* repressor.

Figures 17A and 17B show an example of a promoter for each RNA polymerase.[126,128] Regions underlined indicate sequences highly conserved between var-

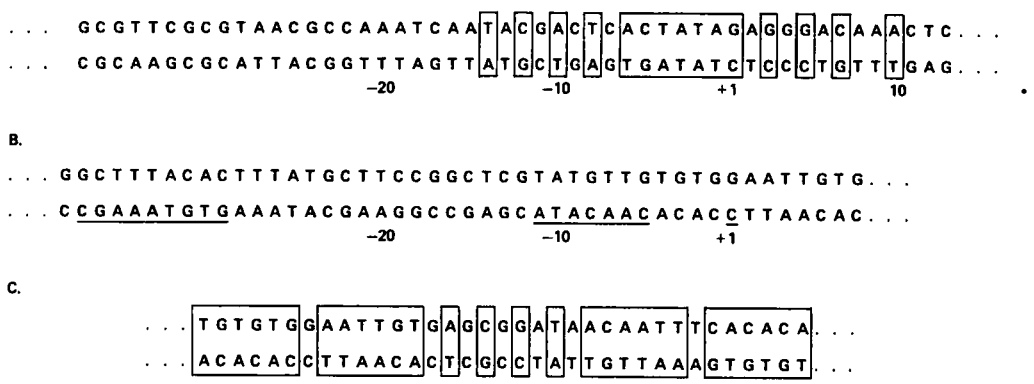


FIGURE 17. (A) Sequence of one of the eight promoters on the T7 genome for the phage-specified RNA polymerase. (B) The wild-type *lac* promoter for *E. coli* RNA polymerase. (C) The *lac* operator sequence.

ious promoters for the same enzyme and thus thought to be important in recognition. Boxed regions indicate symmetrically related sequences.

RNA polymerase binds to the region of DNA called the promoter and initiates an RNA transcript beginning at position +1 of the sequence. This process has a steep temperature dependence, and it has been reasonably proposed that this represents a requirement for melting of the duplex at the promoter, presumably facilitated by its AT-rich nature.[125-129] Unwinding of the template would then expose the coding bases to incoming ribonucleoside triphosphates.

Modification of four tyrosyl residues of the T7 RNA polymerase abolishes promoter binding,[127] suggesting that initiation of promoter melting might involve intercalation of aromatic protein residues with DNA bases, as has been discussed for gene 5P and gene 32P. Also in common with the HDPs, the binding of both RNA polymerases is inhibited at high ionic strength,[125,127] suggesting that ionic interactions play an essential role in complex formation, perhaps by neutralization of backbone phosphate charges with basic residues.

The tetrameric *lac* repressor binds very tightly to the symmetrical *lac* operator sequence shown in Figure 17C.[130,131] It also binds, though less tightly, to any duplex DNA. Complex formation is highly dependent on ionic strength and Record et al.[132] estimate that about eight ion pairs are formed, the release of counterions contributing ~40% of the free energy of the association. A positively charged N-terminal region of *lac* repressor is thought to provide these interactions and be responsible for binding.[133,134] Also in this N-terminal are three tyrosine residues that can be modified by iodination. This modification abolishes operator binding.[135] It is possible that intercalation (or possibly hydrogen bonding) of these residues with DNA bases is involved in specific operator binding.

We see then that in these more complex systems similar mechanisms to those used by the HDPs might be important. The molecular basis of sequence specificity is, however, obscure, although various speculations have been made.[136]

VIII. SUMMARY AND CONCLUSIONS

Helix-destabilizing proteins can be isolated from a great variety of sources including prokaryotic cells and eukaryotic cells, representing a wide phylogenetic range. Specific helix-destabilizing proteins are additionally induced in cells by bacterial and animal

viruses and by transformation. These HDPs vary in size and in quaternary structure in the absence of DNA. All of them bind preferentially and cooperatively to single-stranded DNA, thus promoting denaturation of duplex DNA. They bind without regard to sequence and can saturate all single-stranded regions of the DNA. Many of them also share the ability to promote renaturation of denatured DNA (by holding single-stranded DNA in an extended, favorable conformation) under conditions favoring duplex formation and to stimulate homologous DNA polymerase activity on appropriate templates. In vivo, each appears to have a role in one or more of the processes of DNA replication and its control, recombination, and repair.

The molecular details of DNA binding have been studied extensively in two cases only, gene 5P of coliphage fd and gene 32P of coliphage T4. Intercalation of aromatic residues of the protein with DNA bases appears to be involved in helix-destabilization in both cases. Backbone phosphate charges are neutralized by arrays of basic residues at the DNA binding site of the protein. These interactions are possibly ones commonly used by other proteins binding to DNA.

ACKNOWLEDGMENT

We thank Professor Alex McPherson and colleagues for communicating the results on the crystal structure of gene 5P prior to publication.

REFERENCES

1. **Alberts, B. M., Amodio, F. J., Jenkins, M., Gutman, E. D., and Ferris, F. L.,** Studies with DNA-cellulose chromatography. I. DNA-binding proteins from *Escherichia coli*, *Cold Spring Harbor Symp. Quant. Biol.*, 33, 289, 1962.
2. **Alberts, B. and Herrick, G.,** DNA-cellulose chromatography, *Methods Enzymol.*, 21, 198, 1971.
3. **Alberts, B. and Sternglanz, R.,** Recent excitement in the DNA replication problem, *Nature (London)*, 269, 655, 1977.
4. **Alberts, B. and Frey, L.,** T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA, *Nature (London)*, 227, 1313, 1970.
5. **Snustad, D. P.,** Dominance interactions in *Escherichia coli* cells mixedly infected with bacteriophage T4D wild-type and amber mutants and their possible implications as to type of gene-product function: catalytic vs. stoichiometric, *Virology*, 35, 550, 1968.
6. **Huberman, J. A., Kornberg, A., and Alberts, B.,** Stimulation of T4 bacteriophage DNA polymerase by the protein product of T4 gene 32, *J. Mol. Biol.*, 62, 39, 1971.
7. **Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenburger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, G. H., and Lielausis, A.,** Physiological studies of conditional lethal mutants of bacteriophage T4 D., *Cold Spring Harbor Symp. Quant. Biol.*, 28, 375, 1963.
8. **Warner, H. R. and Hobbs, M. D.,** Incorporation of uracil-^{14}C into nucleic acids in *Escherichia coli* infected with bacteriophage T4 and T4 amber mutants, *Virology*, 33, 376, 1967.
9. **Aposhian, H. V. and Kornberg, A.,** Enzymatic synthesis of DNA. IX. The polymerase formed after T2 phage infection, *J. Biol. Chem.*, 237, 519, 1962.
10. **De Waard, A., Paul, A. V., and Lehman, I. R.,** The structural gene for deoxyribonucleic acid polymerase in bacteriophages T4 and T5, *Proc. Natl. Acad. Sci. U.S.A.*, 54, 1241, 1965.
11. **Warner, H. R. and Barnes, J. E.,** Deoxyribonucleic acid synthesis in *Escherichia coli* infected with some deoxyribonucleic acid polymerase-less mutants of bacteriophage T4, *Virology*, 28, 100, 1966.
12. **Wu, J. -R. and Yeh, Y. -C.,** Requirement of a functional gene 32 product of bacteriophage T4 in UV repair, *J. Virol.*, 12, 758, 1973.
13. **Tomizawa, J., Anraku, N., and Iwama, Y.,** Molecular mechanisms of genetic recombination in bacteriophage. VI. A mutant defective in the joining of DNA molecules, *J. Mol. Biol.*, 21, 247, 1966.

14. **Kozinski, A and Felgenhauer, Z. Z.**, Molecular recombination in T4 bacteriophage DNA. II. Single strand-breaks and exposure of uncomplemented areas as a prerequisite for recombination, *J. Virol.*, 1, 1193, 1967.
15. **Riva, S., Cascino, A., and Geiduschek, E. P.**, Coupling of late transcription to viral replication in bacteriophage T4 development, *J. Mol. Biol.*, 54, 85, 1970.
16. **Carroll, R. B., Neet, K. E., and Goldthwait, D. A.**, Self-association of gene-32 protein of bacteriophage T4, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 2741, 1972.
17. **Carroll, R. B., Neet, K., and Goldthwait, D. A.**, Studies of the self-association of bacteriophage T4 gene 32 protein by equilibrium sedimentation, *J. Mol. Biol.*, 91, 275, 1975.
18. **Kelly, R. C. and von Hippel, P. H.**, DNA "melting" proteins. III. Fluorescence "mapping" of the nucleic acid binding site of bacteriophage T4 gene 32-protein, *J. Biol. Chem.*, 251, 7229, 1976.
19. **Delius, H., Mantell, N. J., and Alberts, B.**, Characterization by electron microscopy of the complex formed between T4 bacteriophage gene 32-protein and DNA, *J. Mol. Biol.*, 67, 341, 1972.
20. **Hosoda, J., Takacs, B., and Brack, C.**, Denaturation of T4 DNA by an *in vitro* processed gene 32-protein, *FEBS Lett.*, 47, 338, 1974.
21. **Anderson, R. A. and Coleman, J. E.**, Physicochemical properties of DNA binding proteins: gene 32 protein of T4 and *E. coli* unwinding protein, *Biochemistry*, 14, 5485, 1975.
22. **Williams, K. R. and Konigsberg, W.**, Structural changes in the T4 gene 32 protein induced by DNA and polynucleotides, *J. Biol. Chem.*, 253, 2463, 1978.
23. **Moise, H. and Hosoda, J.**, T4 gene 32 protein model for control of activity at replication fork, *Nature (London)*, 259, 455, 1976.
24. **Jensen, D. E., Kelly, R. C., and von Hippel, P. H.**, DNA "melting" proteins. II. Effects of bacteriophage T4 gene 32-protein binding on the conformation and stability of nucleic acid structure, *J. Biol. Chem.*, 251, 7215, 1976.
25. **Alberts, B. M.**, Function of the gene 32 protein, a new protein essential for the genetic recombination and replication of T4 phage DNA, *Fed. Proc. Fed. Am. Soc. Biol.*, 29, 1154, 1970.
26. **Marvin, D. A. and Hohn, B.**, Filamentous bacterial viruses, *Bacteriol. Rev.*, 53, 172, 1969.
27. **Berkowitz, S. A. and Day, L. A.**, Molecular weight of single-stranded fd bacteriophage DNA. High speed equilibrium sedimentation and light scattering measurements, *Biochemistry*, 13, 4825, 1974.
28. **Anderson, R. A., Nakashima, Y., and Coleman, J. E.**, Chemical modifications of functional residues of fd gene 5 DNA-binding protein, *Biochemistry*, 14, 907, 1975.
29. **von Hippel, P. H., Jensen, D. E., Kelly, R. C., and McGhee, J. D.**, Molecular approaches to the interaction of nucleic acids with "melting" proteins, in *Nucleic Acid-Protein Recognition*, Vogel, H., Ed., Academic Press, New York, 1977, 65.
30. **Greve, J., Maestre, M. F., Moise, H., and Hosoda, J.**, Circular dichroism study of the interaction between T4 gene 32 protein and polynucleotides, *Biochemistry*, 17, 887, 1978; Circular dichroism of the interaction of a limited hydrolysate of T4 gene 32 protein with T4 DNA and poly[d(A-T)] · poly[(A-T)], *Biochemistry*, 17, 893, 1978.
31. **Coleman, J. E., Anderson, R. A., Ratcliffe, R. G., and Armitage, I. M.**, Structure of gene 5 protein-oligonucleotide complexes as determined by ^{1}H, ^{19}F, and ^{31}P nuclear magnetic resonance, *Biochemistry*, 15, 5419, 1976.
32. **Coleman, J. E. and Armitage, I. M.**, Multinuclear NMR applied to the study of the solution structure of protein-nucleoide complexes, in *NMR in Biology*, Dwek, R. A., Campbell, I. D., Richards, R. E., and Williams, R. J. P., Eds., Academic Press, New York, 1977, 63.
33. **Coleman, J. E. and Armitage, I. M.**, Tyrosyl-base-phenlyalanyl intercalation in gene 5 protein-DNA complexes: ^{1}H NMR of selectively deuterated gene 5 protein, *Biochemistry*, 17, 5038, 1978.
34. **Kelly, R. C., Jensen, D. E., and von Hippel, P. H.**, DNA "melting" protein. IV. Fluorescence measurements of binding parameters for bacteriophage T4 gene 32-protein to mono-, oligo- and polynucleotides, *J. Biol. Chem.*, 251, 7240, 1976.
35. **Poglazov, B. F. and Levshenko, M. T.**, Bacteriophage T4D-induced proteinase, *J. Mol. Biol.*, 84, 463, 1974.
36. **Tsugita, A. and Hosoda, J.**, DNA binding site of the helix destabilizing protein gp 32 from bacteriophage T4, *J. Mol. Biol.*, 122, 255, 1978.
37. **Alberts, B., Barry, J., Bittner, M., Davies, M., Hanna-Inaba, H., Liu, C., Mace, D., Moran, L., Morris, C., Piperno, J., and Sinha, N.**, in *Nucleic Acid-Protein Recognition*, Vogel, H., Ed., Academic Press, New York, 1977, 31.
38. **Lilley, D. M. J., Howarth, O. W., Clark, V. M., Pardon, J. F., and Richards, B. M.**, An investigation of the conformational and self-aggregational processes of histones using ^{1}H and ^{13}C nuclear magnetic resonance, *Biochemistry*, 14, 4590, 1975.
39. **Lilley, D. M. J., Howarth, O. W., Clark, V. M., Pardon, J. F., and Richards, B. M.**, The existence of random coil N-terminal peptides-"tails"-in native histone complexes, *FEBS Letters*, 62, 7, 1976.

40. **Nakashima, Y., Dunker, A. K., Marvin, D. A., and Konigsberg, W.,** The amino acid sequence of a DNA binding protein, the gene 5 product of fd filamentous phage, *FEBS Lett.*, 40, 290, 1974.
41. **Nakashima, Y., Dunker, A. K., Marvin, D. A., and Konigsberg, W.,** The amino acid sequence of DNA binding protein, the gene 5 product of fd filamentous phage. Erratum, *FEBS Lett.*, 43, 125, 1974.
42. **Cuypers, T., van Ouderaa, F. J., and deJong, W. W.,** The amino acid sequence of gene 5 protein of phage M13, *Biochem. Biophys. Res. Commun.*, 59, 557, 1974.
43. **McPherson, A., Jurnak, F., Wang, A., Kolpak, F., Rich, A., and Molineux, I.,** The gene 5 product of fd bacteriophage-structure of a DNA unwinding protein and its complexes with DNA, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
44. **Chou, P. Y. and Fasman, G. D.,** β-turns in proteins, *J. Mol. Biol.*, 115, 135, 1977.
45. **Kornberg, A.,** *DNA Synthesis*, W. H. Freeman, San Francisco, 1974.
46. **Nakashima, Y. and Konigsberg, W.,** Reinvestigation of a region of the fd coat protein sequence, *J. Mol. Biol.*, 88, 598, 1974.
47. **Marvin, D. A. and Wachtel, E. J.,** Structure and assembly of filamentous bacterial viruses, *Nature (London)*, 253, 19, 1975.
48. **Nakashima, Y., Wiseman, R. L., Konigsberg, W., and Marvin, D. A.,** Primary structure and side-chain interactions of PFL filamentous bacterial virus coat protein, *Nature (London)*, 253, 68, 1975.
49. **Nozaki, Y., Chamberlain, B. K., Webster, R. E., and Tanford,C.,** Evidence for a major conformational change of coat protein in assembly of fl bacteriophage, *Nature (London)*, 259, 335, 1976.
50. **Marvin, D. A.,** X-ray diffraction and electron microscope studies on the structure of the small filamentous bacteriophage fd, *J. Mol. Biol.*, 15, 8, 1966.
51. **Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J., and Marvin, F. J.,** Filamentous bacterial viruses. XII. Molecular architecture of the class I (fd, lf, lke) virion, *J. Mol. Biol.*, 88, 581, 1974.
52. **Marvin, D. A., Wiseman, R. L., and Wachtel, E. J.,** Filamentous bacterial viruses. XI. Molecular architecture of the class II (pfl, xf) virion, *J. Mol. Biol.*, 82, 121, 1974.
53. **Nozaki, Y., Reynolds, J. A., and Tanford, C.,** Conformation states of a hydrophobic protein. The coat protein of fd bacteriophage, *Biochemistry*, 17, 1239, 1978.
54. **Chang, C. N., Blobel, G., and Model, P.,** Detection of prokaryotic signal peptidase in an *E. coli* membrane fraction. Endoproteolytic cleavage of nascent fl pre-coat protein, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 361, 1978.
55. **Blobel, G. and Dobberstein, B.,** Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma, *J. Cell. Biol.*, 67, 835, 1975.
56. **Wickner, W.,** Asymmetric orientation of phage M13 coat protein in *E. coli* cytoplasmic membranes and in synthetic lipid vesicles, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 1159, 1976.
57. **Salstrom, J. S. and Pratt, D.,** Role of phage M13 gene 5 in single stranded DNA production, *J. Mol. Biol.*, 61, 489, 1971.
58. **Mazur, B. J. and Model, P.,** Regulation of phage fl single stranded DNA synthesis by a DNA binding protein, *J. Mol. Biol.*, 78, 285, 1973.
59. **Pratt, D., Laws, P., and Griffith, J.,** Complex of bacteriophage M13 single-stranded DNA and gene 5 protein, *J. Mol. Biol.*, 82, 425, 1974.
60. **Geider, K. and Kornberg, A.,** Conversion of M13 viral strand to the double stranded replicative forms by purified proteins, *J. Biol. Chem.*, 249, 3999, 1974.
61. **Staudenbauer, W. L. and Hofschneider, P. H.,** Replicaion of bacteriophage M13. Positive role of gene-5 protein in single strand-DNA synthesis, *Eur. J. Biochem.*, 34, 569, 1973.
62. **Alberts, B., Frey, L., and Delius, H.,** Isolation and characterization of gene 5 protein of filamentous bacterial viruses, *J. Mol. Biol.*, 68, 139, 1972.
63. **Pretorius, H. T., Klein, M., and Day, L. A.,** Gene V protein of fd bacteriophage. Dimer formation and the role of tyrosyl groups in DNA binding, *J. Biol. Chem.*, 250, 9262, 1975.
64. **Cavalieri, S. J., Neet, K. E., and Goldthwait, D. A.,** Gene 5 protein of bacteriophage fd: a dimer which interacts co-operativly with DNA, *J. Mol. Biol.*, 102, 679, 1976.
65. **Rasched, I. and Pohl, F. M.,** Oligonucleotides and the quaternary structure of gene 5 protein from filamentous bacteriophage, *FEBS Letters*, 46, 115, 1974.
66. **Oey, J. L. and Knippers, R.,** Properties of the isolated gene 5 protein of bacteriophage fd, *J. Mol. Biol.*, 68, 125, 1972.
67. **Day, L. A.,** Circular dichroism and ultraviolet absorption of a deoxyribonucleic acid binding protein of filamentous bacteriophage, *Biochemistry*, 12, 5330, 1973.
68. **Glazer, A., Delange, R., and Sigman, D.,** *Chemical Modification of Proteins*, Elsevier, New York, 1975.

69. **Anderson, E., Nakashima, Y., and Konigsberg, W.,** Photo-induced cross-linkage of gene 5 protein and bacteriophage fd DNA, *Int. Symp. Photobiology*, Williamsburg, Va., 1975.
70. **Hull, W. E. and Sykes, B. D.,** Fluorine-19 nuclear magnetic resonance study of fluorotyrosine alkaline phosphatase: the influence of zinc on protein structure and a conformational change induced by phosphate binding, *Biochemistry*, 15, 1535, 1976.
71. **Hull, W. E. and Sykes, B. D.,** Fluorotyrosine alkaline phosphatase: internal mobility of individual tyrosines and the role of chemical shift anisotropy as a ^{19}F nuclear spin relaxation mechanism in proteins, *J. Mol. Biol.*, 98, 121, 1975.
72. **Giessner-Prettre, C. and Pullman, B.,** Ring-current effects in the nmr of nucleic acids: a graphical approach, *Biopolymers*, 15, 2277.
73. **Garssen, G. J., Hilbers, C. W., Schoenmakers, J. G. G., and van Boom, J. H.,** Studies on DNA unwinding. Proton and phosphorus nuclear-magnetic-resonance studies of gene V protein from bacteriophage M13, interacting with d(pC-G-C-G), *Eur. J. Biochem.*, 81, 453, 1977.
74. **Dwek, R. A.,** *Nuclear Magnetic Resonance in Biological Systems: Applications to Enzyme Systems*, Clarendon Press, Oxford, 1973.
75. **Dobson, C. M.,** The structure of lysozyme in solution, in *NMR in Biology*, Dwek, R. A., Campbell, I. D., Richards, R. E., and Williams, R. J. P., Eds., Academic Press, New York, 1977, 63.
76. **Campbell, J. D.,** An NMR view of protein structure in *NMR in Biology*, Dwek, R. A., Campbell, I. D., Richards, R. E., and Williams, R. J. P., Eds., Academic Press, New York, 1977, 33.
77. **McPherson, A.,** Crystallization of proteins from polyethylene glycol, *J. Biol. Chem.*, 251, 6300, 1976.
78. **Sigal, N., Delius, H., Kornberg, T., Gefter, M. L., and Alberts, B.,** A DNA-unwinding protein isolated from *Escherichia coli*: its interaction with DNA and DNA polymerases, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 3537, 1972.
79. **Molineux, J. J., Friedman, S., and Gefter, M. L.,** Purification and properties of the *Escherichia coli* deoxyribonucleic acid-unwinding protein, *J. Biol. Chem.*, 249, 6090, 1974.
80. **Weiner, J. H., Bertsch, L. L., and Kornberg, A.,** The deoxyribonucleic acid unwinding protein of *Escherichia coli*. Properties and functions in replication, *J. Biol. Chem.*, 250, 1972, 1975.
81. **Christiansen, C. and Baldwin, R. L.,** Catalysis of DNA reassociation by the *Escherichia coli* DNA binding protein. A polyamine-dependent reaction, *J. Mol. Biol.*, 115, 441, 1977.
82. **Schaller, H., Uhlmann, A., and Geider, K.,** A DNA fragment from the origin of single-strand to double-strand DNA replication of bacteriophage fd, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 49, 1976.
83. **Geider, K., Beck, E., and Schaller, H.,** An RNA transcribed from DNA at the origin of phage fd single-strand to replicative form conversion, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 645, 1978.
84. **Schekman, R., Weiner, A., and Kornberg, A.,** Multienzyme systems of DNA replication, *Science*, 186, 987, 1974.
85. **Schekman, R., Weiner, J. H., Weiner, A., and Kornberg, A.,** Ten proteins required for the conversion of ϕX174 single stranded DNA to duplex form in vitro, Resolution and reconstitution, *J. Biol. Chem.*, 250, 5859, 1975.
86. **Bouche, J. -P., Rowen, L., and Kornberg, A.,** The RNA primer synthesized by primase to initiae phage DNA G4 replication, *J. Biol. Chem.*, 253, 765, 1978.
87. **Scott, J. F., Eisenberg, S., Bertsch, L. L., and Kornberg, A.,** A mechanism of duplex DNA replication revealed by enzymatic studies of phage ϕX174: catalytic strand separation in advance of replication, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 193, 1977.
88. **Lewin, B.,** *Plasmids and Phages*, Vol. 3, Interscience, New York, 1977, 752.
89. **Grippo, P. and Richardson, C. C.,** DNA polymerase of phage T7, *J. Biol. Chem.*, 246, 6867, 1971.
90. **Reuben, R. C. and Gefter, M. L.,** A DNA-binding protein induced by bacteriophage T7, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 1846, 1973.
91. **Reuben, R. C. and Gefter, M. L.,** A deoxyribonucleic acid-binding protein induced by bacteriophage T7. Purification and properties of the protein, *J. Biol. Chem.*, 249, 3843, 1974.
92. **Scherzinger, E., Litfin, F., and Jost, E.,** Stimulation of T7 DNA polymerase by a new phage-coded protein, *Mol. Gen. Genet.*, 123, 247, 1973.
93. **Hotta, Y. and Stern, H.,** A protein which binds to single-stranded DNA in meiosis, *J. Cell. Biol.*, 47, 92a, 1970.
94. **Hotta, Y. and Stern, H.,** A DNA-binding protein in meiotic cells of *Lilium*, *Dev. Biol.*, 26, 87, 1971.
95. **Hotta, Y. and Stern, H.,** Meiotic protein in spermatocytes of mammals, *Nature (London) New Biol.*, 234, 83, 1971.
96. **Mather, J. and Hotta, Y.,** A phosphorylatable DNA-binding protein associated with a lipoprotein fraction from rat spermatocyte nuclei, *Exp. Cell Res.*, 109, 181, 1977.
97. **Salas, J. and Green, H.,** Proteins binding to DNA and their relation to growth in cultured mammalian cells, *Nature (London) New Biol.*, 229, 165, 1971.

98. Tsai, R. L. and Green, H., Study of intracellular collagen precursors using DNA-cellulose chromatography, *Nature (London) New Biol.*, 237, 171, 1972.
99. Rubio, V., Tsai, W. -P., Rand, K., and Long, C., A comparison of DNA binding proteins from normal and transformed mouse cells, *Int. J. Cancer*, 12, 545, 1973.
100. Rubio, V. and Long, C., Synthesis of a DNA binding protein in normal and transformed cells, *Int. J. Cancer*, 14, 348, 1974.
101. Tsai, R. L. and Green, H., Studies on a mammalian cell protein (P8) with affinity for DNA *in vitro*, *J. Mol. Biol.*, 73, 307, 1973.
102. Fox, T. O. and Pardee, A B., Proteins made in the mammalian cell cycle, *J. Biol. Chem.*, 246, 6159, 1971.
103. Choe, B. -K. and Rose, N. R., Synthesis of DNA-binding proteins during the cell cycle of WI-38 cells, *Exp. Cell Res.*, 83, 271, 1973.
104. Choe, B. -K. and Rose, N. R., Synthesis of DNA-binding protein in WI-38 cells stimulated to synthesize DNA by medium replacement, *Exp. Cell Res.*, 83, 261, 1974.
105. Stein, G. H., DNA binding proteins in young and senescent normal human fibroblasts, *Exp. Cell Res.*, 90, 237, 1975.
106. Van der Vliet, P. C. and Levine, A. J., DNA-binding proteins specific for cells infected by adenovirus, *Nature (London) New Biol.*, 246, 170, 1973.
107. Shanmugam, G., Bhaduri, S., Arens, M., and Green, M., DNA binding proteins in the cytoplasm and in nuclear membrane complex isolated from uninfected and adenovirus 2 infected cells, *Biochemistry*, 14, 332, 1975.
108. Yeh, W. -S.,McGuire, M., Center, M. S., and Consigli, R. A., Partial purification and properties of a DNA-binding protein from nuclei of cells infected with polyoma virus, *B. B. A. Libr.*, 418, 266, 1976.
109. Banks, G. R. and Spanos, A., The isolation and properties of a DNA-unwinding protein from *Ustilago maydis*, *J. Mol. Biol.*, 93, 63, 1975.
110. Donnelly, T. E., Jr., Westergaard, O., and Klenow, H., Isolation and characterization of a DNA-binding non-histone protein from *Tetrahymena pyriformis*, *B. B. A. Libr.*, 402, 150, 1975.
111. Herrick, G. and Alberts, B., Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus, *J. Biol. Chem.*, 251, 2124, 1976.
112. Herrick, G. and Alberts, B., Nucleic acid helix-coil transitions mediated by helix-unwinding proteins from calf thymus, *J. Biol. Chem.*, 251, 2133, 1976.
113. Herrick, G., Delius, H., and Alberts, B., Single-stranded DNA structure and DNA polymerase activity in the presence of nucleic acid helix-unwinding proteins from calf thymus, *J. Biol. Chem.*, 251, 2142, 1976.
114. Thomas, T. L. and Patel, G. L., DNA unwinding component of the nonhistone chromatin proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 4364, 1976.
115. Duguet, M. and deRecondo, A. -M., A deoxyribonucleic acid unwinding protein isolated from regenerating rat liver. Physical and functional properties, *J. Biol. Chem.*, 253, 1660, 1978.
116. Otto, B., Baynes, M., and Knippers, R., A single-strand-specific DNA-binding protein from mouse cells that stimulates DNA polymerase. Its modification by phosphorylation, *Eur. J. Biochem.*, 73, 17, 1977.
117. Sen, A. and Todaro, G. J., Species-specific cellular DNA-binding proteins expressed in mouse cells transformed by chemical carcinogens, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 1647, 1978.
118. Huang, A. T. -F., Riddle, M. M., and Koons, L. S., Some properties of a DNA-unwinding protein unique to lymphocytes from chronic lymphocytic leukemia, *Cancer Res.*, 35, 981, 1975.
119. Kornberg, R. D., Structure of chromatin, *Annu. Rev. Biochem.*, 46, 931, 1977.
120. Stein, G. S. and Kleinsmith, L. J., *Chromosomal Proteins and their Role in the Regulation of Gene Expression*, Academic Press, New York, 1975.
121. Wang, J. C., Interaction between DNA and *Escherichia coli* protein ω, *J. Mol. Biol.*, 55, 523, 1971.
122. Gellert, M., Mizuuchi, K., O'Dea, M. H., and Nash, H. A., DNA gyrase: an enzyme that introduces superhelical turns into DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3872, 1976.
123. Liu, L. F. and Wang, J. C., *Micrococcus luteus* DNA gyrase: active components and a model for its supercoiling of DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 2098, 1978.
124. Roberts, R. J., Restriction endonucleases, *CRC Crit. Rev. Biochem.*, 4(2), 123, 1976.
125. Chamberlin, M. J., Interaction of RNA polymerase with the DNA template, in *RNA Polymerase*, Losick, R. and Chamberlin, M., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y., 1976, 159.
126. Gilbert, W., Starting and stopping sequences for the RNA polymerase, in *RNA Polymerase*, Losick, R. and Chamberlin, M., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y., 1976, 193.

127. **Oakley, J. L., Pascale, J. A., and Coleman, J. E.,** T7 RNA polymerase: conformation, functional groups, and promoter binding, *Biochemistry*, 14, 4684, 1975.
128. **Oakley, J. L. and Coleman, J. E.,** Structure of a promoter for T7 RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 4266, 1977.
129. **Oakley, J. L., Strothkamp, R. E., Sarris, A. H., and Coleman, J. E.,** T7 RNA polymerase: promoter structure and polymerase binding, *Biochemistry*, 18, 528, 1979.
130. **Maxam, A. and Gilbert, W.,** The nucleotide sequence of the *lac* operator, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 3581, 1973.
131. **Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S.,** Genetic regulation: the *lac* control region, *Science*, 187, 27, 1975.
132. **Record, T. M., Jr., de Haseth, P. L., and Lohman, T. M.,** Interpretation of monovalent and divalent cation effects on the *lac* repressor-operator interaction, *Biochemsitry*, 16, 4791, 1977.
133. **Beyreuther, K., Adler, K., Geisler, N., and Klemm, A.,** The amino-acid sequence of *lac* repressor, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 3576, 1973.
134. **Platt, T., Files, J. G., and Weber, K.,** *Lac* repressor. Specific proteolytic destruction of the NH_2-terminal region and loss of the deoxyribonucleic acid-binding activity, *J. Biol. Chem.*, 248, 110, 1973.
135. **Fanning, T. G.,** Iodination of *Escherichia coli lac* repressor. Effect of tyrosine modification on repressor activity, *Biochemistry*, 14, 2512, 1975.
136. **Seeman, N. C., Rosenberg, J. M., and Rich, M.,** Sequence-specific recognition of double helical nucleic acids by proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 804, 1976.
137. **Coleman, J. E.,** unpublished data.
138. **Chou, P. Y. and Fasman, G. D.,** Prediction of protein conformation, *Biochemistry*, 13, 222, 1974.