Journal of Supramolecular Structure 10:457–465 (1979)
Biological Recognition and Assembly 15–23

Structure of the DNA Binding Cleft of the Gene 5 Protein From Bacteriophage fd

Alexander McPherson, Frances Jurnak, Andrew Wang, Ian Molineux, and Alexander Rich

Department of Biological Chemistry, Milton S. Hershey Medical Center of The Pennsylvania State University, Hershey, Pennsylvania 17033 (A.M., F.J.), Massachusetts Institute of Technology, Cambridge, Massachusetts 01239 (A.W., A.R.), and Department of Microbiology, University of Texas, Austin, Texas 78712 (I.M.)

The structure of the gene 5 DNA unwinding protein from bacteriophage fd has been solved to 2.3-Å resolution by X-ray diffraction techniques. The molecule contains an extensive cleft region that we have identified as the DNA binding site on the basis of the residues that comprise its surface. The interior of the groove has a rather large number of basic amino acid residues that serve to draw the polynucleotide backbone into the cleft. Arrayed along the external edges of the groove are a number of aromatic amino acid side groups that are in position to stack upon the bases of the DNA and fix it in place. The structure and binding mechanism as we visualize it appear to be fully consistent with evidence provided by physical-chemical studies of the protein in solution.

Key words: DNA binding protein, gene 5, fd bacteriophage, X-ray diffraction, protein-nucleic acid interactions

The gene 5 product of the filamentous bacteriophage fd is a single-strand specific DNA binding protein of 10,000 daltons molecular weight made in approximately 100,000 copies per infected E coli cell [1]. The protein is coded by the phage genome and is elaborated late in infection when the transition from double-stranded replicative form DNA to single-stranded synthesis of the daughter viral genomes occurs [2]. Its primary physiologic role is the stabilization and protection of single-strand DNA daughter virions from duplex formation following replication in the host [3]. Under conditions of low ionic strength in vitro, it will melt double-stranded homopolymers and will reduce the melting temperature of native double-strand calf thymus DNA by 40°C [4].

Received April 16, 1979; accepted April 18, 1979.

0091-7419/79/1004-0457$02.00

Gene 5 protein exists predominantly as a dimer when free in solution [5] and binds, with a stoichiometry of one monomer per four bases [1], to DNA chains running in opposite directions so that it cross-links two strands of a duplex or opposite sides of closed circular single-stranded DNA. The mechanism for DNA unwinding is simply a linear aggregation along the two opposite strands and derives from the highly cooperative nature of the lateral binding interactions [6]. The extensive degree of cooperativity is presumably a product of strong protein-protein forces between adjacent molecules of the gene 5 protein along the DNA strands. On binding to circular, single-stranded fd DNA, the gene 5 protein collapses the circle into a helical rod-like structure containing two anti-parallel strands of DNA [1].

We have solved the structure of the gene 5 protein to 2.3-Å resolution, traced the course of the polypeptide backbone, and constructed a Kendrew model of the molecule that includes all nonhydrogen atoms in the structure. The technical details of the analysis will be presented elsewhere [7], but we will describe here some of the features of the structure directly relevant to the means and interactions by which the single-stranded DNA is bound to the protein. We would like, in addition, to compare some of our findings with those derived from physical and chemical studies of the gene 5 protein in solution.

MATERIALS AND METHODS

The gene 5 protein was isolated from phage lysates and crystallized for X-ray diffraction analysis as described by McPherson et al [8]. Diffractometer data on native crystals plus six isomorphous heavy-atom derivatives were used to calculate to 2.3-Å resolution an electron density map of the gene 5 protein. The phases were deduced using the isomorphous replacement technique of Perutz [9] and refined by the procedures of Dickerson et al [10] by means of the error treatment of Blow and Crick [11]. The sequence of the protein, as determined by Nakashima et al [12], was fitted to the electron density map and a model was then constructed on a scale of 2 cm/Å in a Richards optical comparator [13] using Kendrew model parts. The coordinates from the model were obtained for all atoms using a plumb and line. Graphical presentation and illustration were produced using the program VDW written by Ringle and Hanson and modified by P.M.D. Fitzgerald for the PDP 11/40 computer. Although not yet complete, the atomic coordinates are presently being refined using the constrained least-squares procedures of Konnert and Hendrickson. Many of our deductions regarding the binding of single-stranded DNA to the gene 5 protein are derived from our attempts to fit by hand a model tetranucleotide to the putative binding region of the gene 5 protein Kendrew model. We will pursue this approach using an interactive graphics system when model refinement has been completed.

RESULTS

Figure 1 shows a wooden model of the gene 5 protein at an effective resolution of about 5.0 Å, viewed approximately down the crystallographic 011 direction. The monomer is roughly 45 Å long, 25 Å wide, and 30 Å high. It is essentially globular, with an appendage of density closely approaching the molecular dyad and tightly interlocking with an identical symmetry-related appendage on the second molecule within the dimer. The major portion of the molecular density slants from upper left to lower right in Figure 1, and creates an overhanging ledge of density that serves in part to create an extended shallow groove banding the outside waist of the monomer. In the dimer the two symmetry-related

grooves, each about 30 Å in length, run antiparallel courses and are separated by roughly 25 Å.

The course of the polypeptide chain in the gene 5 monomer is shown in Figure 2 as deduced from our 2.3-Å electron density map. The protein is composed entirely of antiparallel β structure with no α-helix whatsoever. This is as expected from spectroscopic measurements [14] and sequence-structure rules [15]. There are three basic elements of secondary structure shown in Figure 3 that comprise the molecule, a three-stranded antiparallel β sheet arising from residues 12–49, a two-stranded antiparallel β ribbon formed by residues 50–70, and a second two-stranded antiparallel β ribbon derived from residues 71–82. It is the first of the two β loops (50–70) that creates the appendage of density near the molecular dyad and maintains the dimer in solution. The second β loop (71–82) forms the top surface of the molecule and we believe it is most involved in producing the neighbor-neighbor interactions responsible for the cooperative protein binding. The central density of the molecule is created by the severely twisted three-stranded β sheet made up of residues 12–49. As a result of the distortion from planarity of these three strands, a distinct concavity is produced on the underside of this sheet. Enhanced in part by density from the β ribbon (50–70) near the dyad, this concavity is extended and deepened to provide the long 30-Å groove.

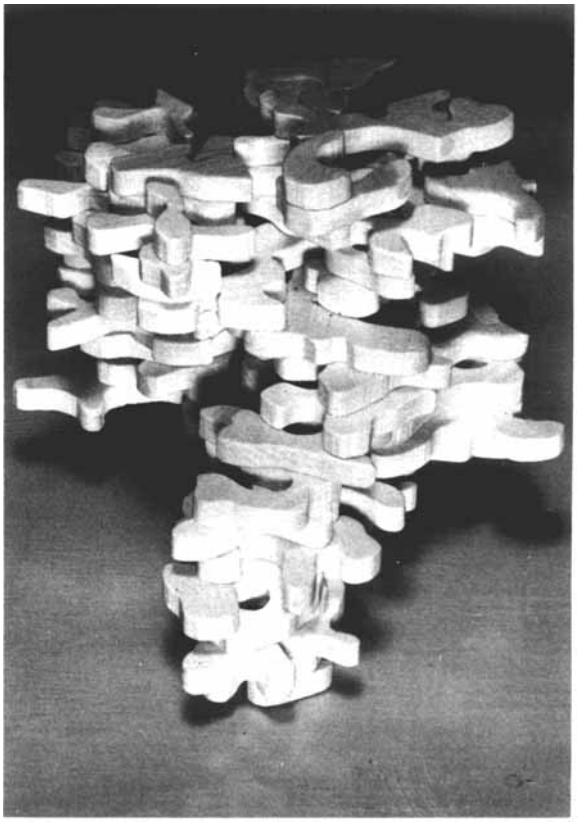

Fig. 1. Representation of the gene 5 protein electron density based on the 2.3-Å Fourier made by cutting appropriate envelopes of density from each section of map and assembling them in the y direction. Model is viewed approximately along the 011 direction and can be seen as an essentially globular mass with a protrusion of density near the molecular dyad.

The long groove beneath the three-stranded sheet by its shape and extent seems to be the single-strand DNA binding interface. There is no other passage through the density that would be consistent with a long polynucleotide binding region. Given that this is the site, then the mode of cross-strand attachment of the gene 5 protein would be that shown in Figure 4. The two monomers within the dimer bind to strands of opposite polarity across the duplex DNA with the molecular dyad roughly perpendicular to the plane of the two bound strands which are separated in the complex by about 25 Å.

Considering that refinement of the structure is not yet complete and that placement of amino acid side chains is still tentative, there still remain a number of interesting features that can be described and that are not likely to be seriously revised. The tetranucleotide binding trough in the gene 5 protein is composed primarily of the amino acid side chains arising from residues 12–49 of the antiparallel β sheet. These strands run more or less parallel with the direction of the DNA strand as it would bind in the trough. The surface of the trough is also comprised in part of residues 50–56 and 66–69, from the interior portions of the two strands forming the β loop near the molecular dyad. A stereo drawing of the gene 5 monomer showing the binding region is shown in Figures 5 and 6.

Aromatic amino acid side chains have been implicated in the binding of DNA to the gene 5 protein by chemical modification and neutron magnetic resonance (NMR) studies. These show that tyrosines 26, 41, and 56 lie near the surface of the protein and are readily nitrated by tetranitromethane, which prevents DNA binding [15]. Conversely, binding of oligonucleotides or DNA prior to reaction prevents nitration of these residues. ^{19}F NMR of the fluorotyrosyl-containing protein confirms these results and further suggests that these tyrosyls intercalate or stack with the bases of the DNA [16]. Similar results, obtained with deuterated protein, implicate at least one phenylalanine residue in a similar fashion [17]. Spectral data lend further support to the idea that the aromatic residues of the protein stack upon or intercalate between the bases [18].

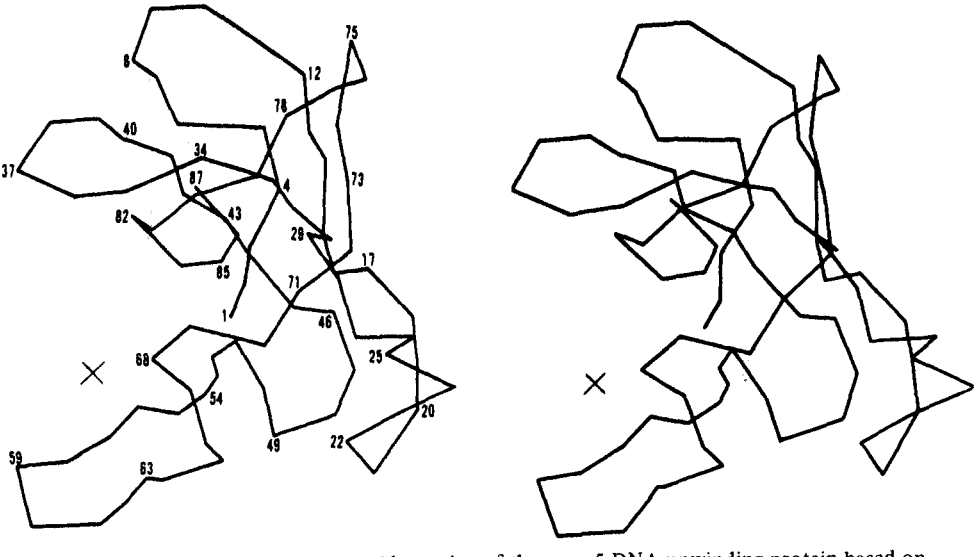

Fig. 2. Stereo drawing of the polypeptide tracing of the gene 5 DNA unwinding protein based on α-carbon coordinates measured from a Kendrew model. The view is roughly along the crystallographic *a* axis.

A number of aromatic residues are arrayed along the binding surface, and these include tyrosines 26, 41, 34, and 56 and phenylalanines 13 and 68. The distribution is not uniform, one end of the trough appearing considerably richer than the other and bearing both phenylalanines as well as tyrosines 34, 41, and 56. The opposite end of the trough, that closest to the viewer in Figure 6, contains only tyrosine 26. The aromatic side chains, with the exception of tyrosine 56 and phenylalanine 68, do not protrude into the binding cleft but are turned away. Each can, however, be brought down into the binding groove by an appropriate rotation about the β carbon. Of particular interest are the side groups of tyrosine 41, tyrosine 34, and phenylalanine 13, which form a triple stack with phe 13 most interior, tyr 41 fully on the outside, and tyr 34 sandwiched in between. The stacking is not precisely one atop the other, but the rings are fanned out like three playing cards. These rings are on the upper edge of the trough; below them on the lower edge and actually positioned in the mouth of the groove is tyrosine 56. Coleman et al [16] note from their NMR data that in the uncomplexed protein a number of tyrosyl proton resonances show upfield shifts, suggesting some ring current effects due to stacking. They hypothesized that the tyrosyl residues involved might be in some organized array such as we observe. These resonances are lost on oligonucleotide binding, suggesting a disruption of the pattern as the residues begin interacting with the bases of the DNA.

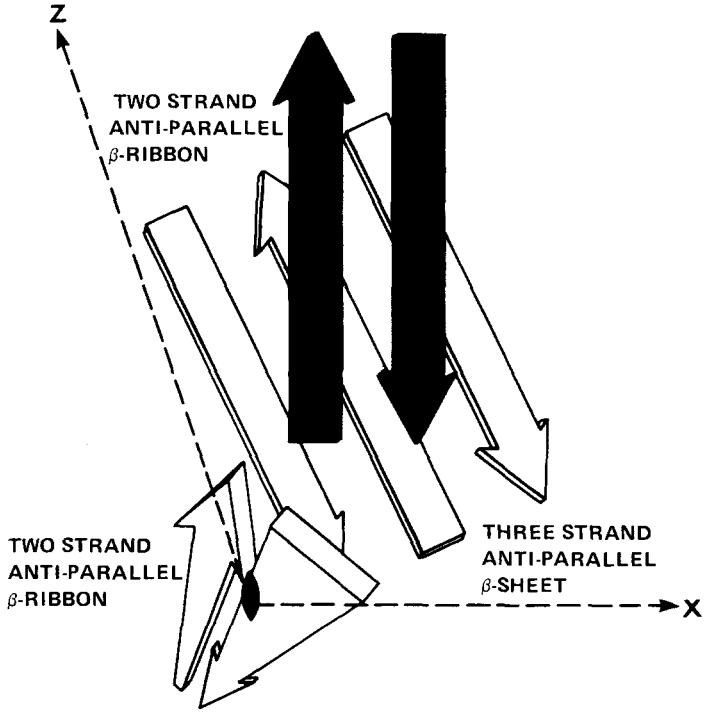

Fig. 3. Schematic drawing showing the secondary structural elements of the gene 5 molecule. Beginning with the N-terminal portion of the sequence, it consists of a three-stranded β sheet which forms the major part of the DNA binding region, a two-stranded β ribbon principally responsible for maintaining the molecule as a dimer in solution, and a second β ribbon that we believe is the primary participant in the lateral interactions from which the cooperativity of protein binding arises.

Tyrosine 26 is near the β bend between strands 1 and 2 of the antiparallel β sheet. This bend appears to be a very flexible elbow of density extending out from the central mass of the molecule and making up one end of the binding region. Even in the crystal, it projects into a large solvent area and seems to be rather mobile and free to move. It is the only tyrosine that we were able to iodinate in the crystal.

We noted that three of the tyrosines in the molecule – 26, 41, and 56 – fall adjacent to, or one removed from, a proline residue. The backbone structure of the protein is engaged in β structure and one would expect that this hydrogen bonding network might restrict the freedom of many bulky side groups. However, by virtue of their proximity to a natural structure-disrupting amino acid, proline, these three tyrosines are endowed with more freedom than they might otherwise enjoy. Because of the proline residues, the tyrosine side chains can rotate from one side of the sheet to the other through the trap door created by the neighbor.

Cysteine 33 is on the inside surface of the binding groove and could certainly interact with the DNA strand. In the conformation that we observe, however, the –SH group is turned up into the interior of the molecule, away from the solvent. It is not in contact with the neighboring tyrosine 34. Although inaccessible to the bulkier Ellmans' reagent, the single cysteine can be reacted with mercuric chloride. Mercuration of cysteine 33 prevents nucleotide or DNA binding to the protein and, conversely, complexation with oligonucleotides prevents reaction with mercury [15]. This is consistent with its placement in the binding groove as is the finding that this –SH group can be photo-cross-linked to thymidine residues of bound nucleic acid [19].

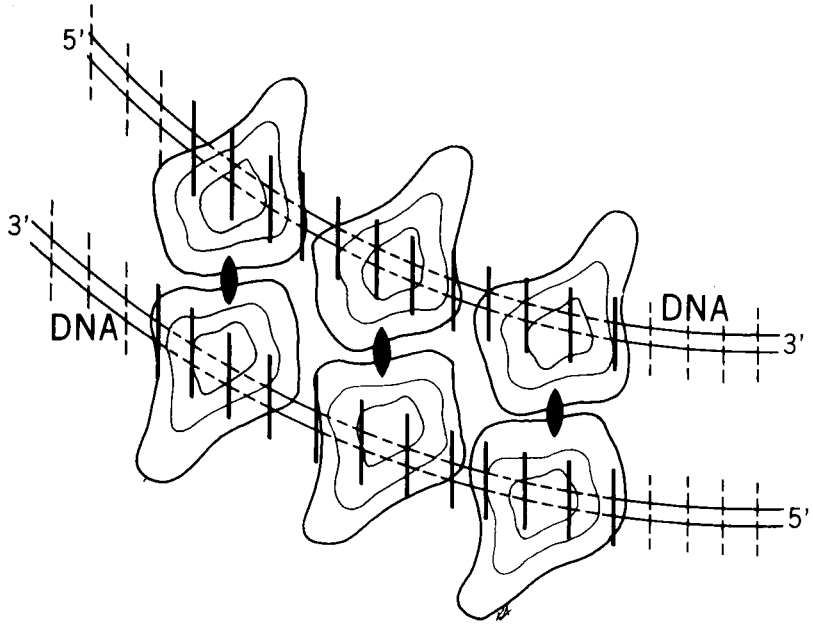

Fig. 4. Schematic representation illustrating the cross-chain binding of the gene 5 dimers to opposing strands of a DNA duplex or opposite sides of a circular single-stranded DNA molecule. The distance between opposing DNA single strands would be about 25 Å.

Acetylation of the ϵ-amino groups of the seven lysyl residues destroys the binding of gene 5 protein to oligonucleotides and DNA, but these groups are not protected by the presence of DNA from reaction [15]. In addition, NMR spectra show that the ϵ-amino groups do not undergo chemical shift or line broadening upon complexation and appear to remain highly mobile. This was interpreted as indicating that the ϵ-amino groups provide a neutralizing charge cloud for the negative phosphate backbone of the nucleotide but do not form highly rigid salt bridges or hydrogen bonds [16]. Resonances from the δ-CH_2 groups of the arginyl residues do undergo chemical shifts and line broadening on DNA complexation, and this could represent direct interaction of the guanidino groups with the phosphate backbone [16].

The DNA binding trough has over its interior surface a fairly large number of basic amino acid side chains which, because of the length and flexibility of these residues, reach into the groove though originating at disparate locations within the molecule. The basic residues most clearly apparent in the cleft are arginines 21, 80, and 82 and lysines 24 and 46. These are all found on the interior surface of the trough, so that the cleft is also something of a postively charged pocket in the protein. It should be noted that other basic amino acids could conceivably approach the binding region but in the conformation we observe in the crystal they are elsewhere. In particular, arginine 16 and lysine 46 are certainly close to the interface, but we see them turned away from the groove rather than toward it.

DISCUSSION

The DNA binding cleft of about 30-Å length and formed principally by the underside of an antiparallel β sheet has been tentatively identified in the gene 5 protein. This assignment is based on the general shape and size of the groove and the distribution of amino acid residues on its surface, which have been implicated in DNA binding by NMR, optical spectra and chemical modification studies in solution.

The binding cleft is very interesting in that the positively charged residues of lysine and arginine are distributed predominantly over the most interior surface, while the aromatic

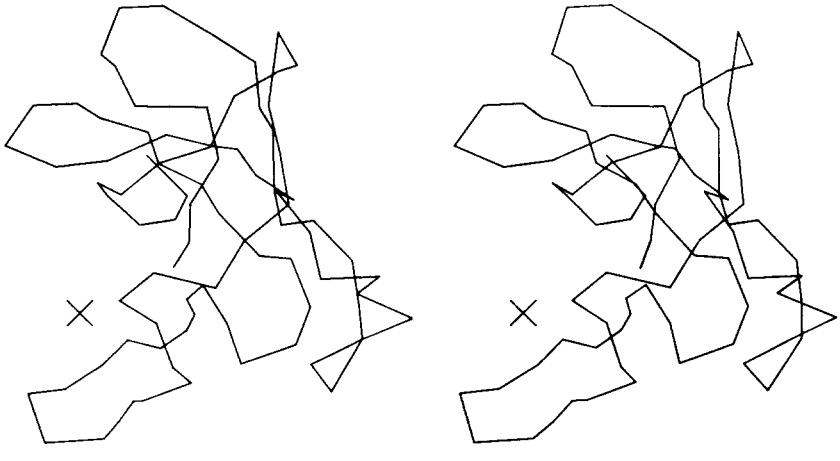

Fig. 5. Stereo representation of the polypeptide backbone of the gene 5 protein rotated so that the view is roughly along the course of the DNA binding groove. This groove is approximately 25 Å in length and runs more or less parallel with the strands of the β sheet.

residues are arrayed primarily along the exterior edges. Thus it appears that the negatively charged polyphosphate backbone of the single-stranded DNA is first recognized by the protein and that it is drawn and fixed to the interior of the groove by the charge interactions. This is followed by rotation of the aromatic groups down and into position to stack upon the bases of the DNA, which are now splayed out toward the exterior of the protein. This is consistent with the finding of Day [14] from micrographic and spectral data that the DNA in the gene 5 complex is completely unstacked and stretched along the filament axis. The cleft then acts as an elongated pair of jaws and the DNA is drawn between them by charge interactions involving the phosphates with the interior lysines and arginines. The jaws then close around the DNA strand through small conformation changes and the rotation of aromatic side chains into position to stack upon the purine and pyrimidine bases. That small, but not gross, conformation changes occur in the protein upon DNA binding is in agreement with the NMR studies by Coleman et al [16] on α-CH and on aliphatic methyl groups which suggest that gene 5 protein must contain a large percentage of fixed structure without large regions of flexible polypeptide chain. Days' spectral evidence also indicates that only small changes occur in overall protein structure on binding [14].

That the interaction between gene 5 protein and DNA is to a great extent electrostatic is clear from the finding that moderate divalent and monovalent cation concentrations cause the complex to dissociate and that binding capacity is lost when the arginines and lysines are chemically modified [15]. The involvement of the aromatic groups, however, is also quite clear from the NMR and spectral data. The minor conformation changes in the gene 5 protein involving other residues and possibly even main-chain atoms are consistent with the physical and chemical studies. Therefore, although our binding mechanism is somewhat speculative, it is to our knowledge entirely consistent with the structure as we visualize it and with the evidence at hand from noncrystallographic analyses.

The binding of the gene 5 DNA unwinding protein to deoxyoligonucleotides is nonspecific in that complexation will occur with oligomers of any sequence. It is found, however, that the gene 5 protein binds oligomers of different sequence with differing affinities and these may vary over two orders of magnitude [16]. Thus the protein does distinguish

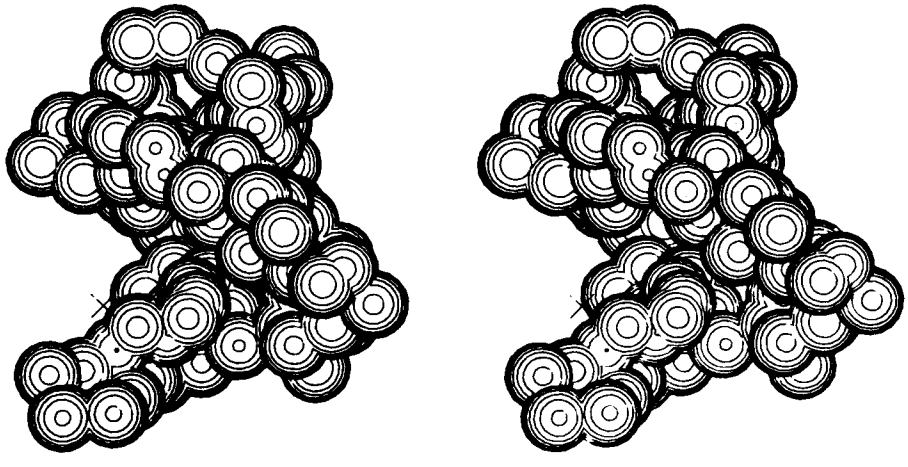

Fig. 6. Stereo drawing of the polypeptide backbone of the gene 5 protein oriented as in Figure 5, with each α-carbon represented by a sphere of 3.0-Å radius to give a space-filling effect. The DNA binding region is the pronounced groove running roughly perpendicular to the plane of the drawing.

between different binding possibilities, and this could be the basis for recognition of specific nucleation sites on the fd DNA. Until the structure of a gene 5 protein oligomer complex has been directly visualized we will not be able to confidently establish the interactions that confer the differential binding affinities. The possibility exists, however, that when the interactions of the gene 5 protein with specific sequences of DNA are completely defined, they will suggest how, by only minor structural alterations, a high degree of recognition specificity might be achieved.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health, National Science Foundation, National Aeronautics and Space Administration, and the American Cancer Society. A.W. is supported by a grant from the MIT Cancer Center. F.J. is an NIH postdoctoral fellow.

REFERENCES

1. Alberts B, Frey L, Delius H: J Mol Biol 68:139–152, 1972.
2. Oey JL, Knippers R: J Mol Biol 68:125–128, 1972.
3. Mazur BJ, Model P: J Mol Biol 78:285–300, 1973.
4. Salstrom JS, Pratt D: J Mol Biol 61:489–501, 1971.
5. Cavalieri S, Goldthwait DA, Neet K: J Mol Biol 102:713, 1976.
6. Dunker AK, Anderson EA: Biochim Biophys Acta 402:31–34, 1975.
7. McPherson A, Jurnak FA, Wang AHJ, Molineux I, Rich A: J Mol Biol (In press).
8. McPherson A, Molineux I, Rich A: J Mol Biol 106:1077, 1976.
9. Bragg WL, Perutz MF: Proc Roy Soc A225:315, 1954.
10. Dickerson RE, Kendrew JC, Strandberg BE: Acta Crystallogr 14:1188, 1961.
11. Blow DM, Crick FHC: Acta Crystallogr 12:794, 1959.
12. Nakashima Y, Dunker AK, Marion DA, Konigsberg W: FEBS Lett 40:290, 1974.
13. Richards FM: J Mol Biol 37:225, 1968.
14. Day LA: Biochemistry 12:5329, 1973.
15. Anderson RA, Nakashima Y, Coleman JE: Biochemistry 14:907, 1975.
16. Coleman JE, Anderson RA, Ratcliffe RG, Armitage IM: Biochemistry 15:5419, 1976.
17. Coleman JE, Armitage IM: Biochemistry (In press).
18. Pretorius HT, Klein M, Day LA: J Biol Chem 250:9262–9269, 1975.
19. Nakashima Y, Konigsberg W: Paper presented at the International Symposium for Photobiology, Williamsburg, Virginia, 1975.