THE STRUCTURE OF A DNA UNWINDING PROTEIN AND ITS COMPLEXES WITH OLIGODEOXYNUCLEOTIDES BY X-RAY DIFFRACTION

Alexander McPherson and Frances Jurnak, *Department of Biochemistry, University of California, Riverside, California 92521*
Andrew Wang, Frank Kolpak, and Alexander Rich, *Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
Ian Molineux, *Department of Microbiology, University of Texas, Austin, Texas 78712*
Paula Fitzgerald, *The Buffalo Medical Foundation, Buffalo, New York 14203 U.S.A.*

ABSTRACT 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 cleft then acts as an elongated pair of jaws that draws the DNA between them by charge interactions involving the phosphates with the interior lysines and arginines. The jaws then close on the DNA strand through small conformation changes and the rotation of aromatic side-chains into position to stack upon the purines and pyrimidines. Complexes of the gene 5 protein with a variety of oligodeoxynucleotides have been formed and crystallized for x-ray diffraction analysis. The crystallographic parameters of four different unit cells indicate that the fundamental unit of the complex is composed of six gene 5 protein dimers. We believe this aggregate has 622 point group symmetry and is a ring formed by end to end closure of a linear array of six dimers. From our results we have proposed a double helical model for the gene 5 protein-DNA complex in which the protein forms a spindle or core around which the DNA is spooled. 5.0-Å x-ray diffraction data from one of the crystalline complexes is currently being analyzed by molecular replacement techniques to obtain what we believe will be the first direct visualization of a protein-deoxyribonucleic acid complex approaching atomic resolution.

INTRODUCTION

Determination of the structure of a complex between a DNA binding protein and fragments of nucleic acid by x-ray diffraction analysis promises to lend considerable insight into the means by which these two important macromolecules interact. In addition to delineating the atomic interactions by which they recognize and bind to one another, knowledge of such a structure could clarify some of the mechanisms by which the flow of genetic information is controlled. In the case of the DNA unwinding protein which we describe here, we believe information may also be gained concerning the assembly and general architectural features of large protein-nucleic acid structures such as are found in chromosomal material and viruses.

The gene 5 product of the filamentous bacteriophage fd is a single strand specific DNA binding protein of 10,000 mol wt having a known sequence (1) and made in ~100,000 copies per infected *Escherichia coli* cell (2). 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 (3). Its primary physiological role is the stabilization and protection of single strand DNA daughter virions from duplex formation after replication in the host (4). Under low ionic strength conditions in vitro, it will melt double stranded homopolymers and will reduce the melting temperature of native double strand calf thymus DNA by 40°C (5).

Gene 5 protein exists predominantly as a dimer when free in solution (6) and binds, with a stoichiometry of one monomer per four bases (2), to DNA chains running in opposite directions so that it crosslinks 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 opposing strands and derives from the highly cooperative nature of the lateral binding interactions (7). The extensive degree of cooperativity is presumably a product of strong protein-protein forces between adjacent molecules of the gene 5 along the DNA strands. On binding to circular single stranded fd DNA, the gene 5 protein collapses the circle into a helical rodlike structure containing two antiparallel strands of DNA (2).

The gene 5 protein-DNA complexes produced in vitro as visualized by electron microscopy are unique in that two protein covered strands coalesce to yield a helical rodlike structure in which there are 12 gene 5 monomers per turn of the helix. The helix has a width of ~100 Å (2). The gene 5 protein-DNA complexes resemble mature filamentous bacteriophage virions though there are clear differences. The mature virus is formed by the displacement of the gene 5 protein at, or in, the host cell membrane by the coat protein, the product of gene 8 (8). The gene 5 protein is never found in the virion but is returned to the cell for reuse.

In vitro complexes of the gene 5 protein with fd phage DNA have been reported to differ in structure from complexes isolated directly from infected cells. These in vivo complexes were observed to be composed of fibers 40 Å in width that were supercoiled to give an overall width of 160 Å and a longitudinal repeat of 160 Å (9). More recent electron microscopy studies by Gray (10), however, find the in vivo and in vitro complexes to be identical and to resemble the helical rods described above. One difference has been noted between the two complexes: the stoichiometry of binding from presumably saturated in vitro complexes is one gene 5 monomer per four nucleotides, while the in vivo complexes tend to give nonintegral values of ~4.6 nucleotides per gene 5 monomer (11).

There is evidence from crosslinking studies in solution that when gene 5 protein is combined with deoxyoligonucleotides from four to eight in length, high molecular weight aggregates containing up to about eight monomers are formed and can be seen on sodium dodecyl sulfate (SDS)-polyacrylamide gels. It was concluded in these studies that the oligomers gave rise to crosslinked aggregates very similar to those obtained with poly (dA-dT) and that the binding of short stretches of nucleic acid chain appears to induce the association of gene 5 monomers to one another (12).

The structure of the gene 5 protein has now been solved to 2.3 Å resolution using conventional isomorphous replacement x-ray diffraction techniques (13, 14). We have traced the course of the polypeptide backbone and constructed a Kendrew model of the molecule that includes all nonhydrogen atoms in the structure. In addition, we have formed complexes of the gene 5 protein with a number of different homogeneous deoxyoligonucleotides in solution and have crystallized a variety of these complexes in a number of crystal forms. It is our intention to use the structure of the native protein to determine the structure of the single crystals of protein-DNA complexes.

CRYSTALLOGRAPHIC ANALYSIS

The gene 5 protein was prepared from fd bacteriophage-infected *E. coli* strain K12 that were harvested 2 h after infection by the methods of Alberts et al. (2). These methods included sequential DNA cellulose and DEAE cellulose chromatography. Homogeneity was confirmed by SDS polyacrylamide gel electrophoresis. Crystallization was achieved using the vapor diffusion technique in glass depression plates. 10 μl of a 15 mg/ml protein solution containing 0.01 M Tris-HCl at pH 7.6 was combined with 10 μl of a 10% PEG 4000 solution and allowed to equilibrate with a 12% PEG 4000 reservoir at 4°C.

The crystals used in the analysis were of monoclinic space group C2 with a = 76.5 Å, b = 28.0 Å, and c = 42.5 Å with β = 108°. There was one monomer of gene 5 in the crystallographic asymmetric unit implying the dimer to have dihedral symmetry. The crystals diffract to at least 1.2 Å resolution and show very little radiation damage at ~100 h of exposure time.

Using the step scan mode (15), Friedel pairs were collected at $\pm 2\theta$ for native and all isomorphous derivative crystals to 2.3-Å resolution. The reflections were recorded on a Picker FACS −1 diffractometer (Picker Corp., Cleveland, Ohio) with a 1,600 w fine focus x-ray tube. A complete data set was comprised of 3,700 independent reflections and could be obtained from one or at most two crystals. In addition, each data set was collected two times on different crystals and averaged with merging residuals of no more than 3.5% on $|\bar{F}|$. Scaling of derivative to native structure amplitudes was carried out in shells of $\sin\theta/\lambda$ and the residuals varied from 16% to 22% on $|\bar{F}|$. Only the iodine derivative appeared non isomorphous beyond 3.0 Å resolution.

The structure of the gene 5 protein was solved to a resolution of 2.3 Å using conventional isomorphous replacement x-ray diffraction techniques. The initial derivative substitution sites, those for $PtBr_2(NH_3)_2$, were located by standard and anomalous difference Patterson syntheses and phases based on these positions were calculated after inclusion of anomalous dispersion data. All of the subsequent derivatives were located by difference Fourier synthesis and backchecked against their corresponding difference Pattersons. The validity of positions was further confirmed by cross-derivative Patterson syntheses and by least squares refinement with sequential omission of each derivative from the phase calculations.

The heavy atom parameters were refined by alternate cycles of phase calculations and least

TABLE I
NATIVE TO DERIVATIVE SCALING STATISTICS

Compound	Total No. reflections	Residual after fitting to native													
		Overall	6.63	4.60	3.87	3.46	3.19	2.98	2.82	2.69	2.58	2.49	2.40	2.33	Beyond
$PTBr_2(Nh_3)_2$	3757	0.208	0.301	0.172	0.172	0.169	0.171	0.178	0.204	0.214	0.249	0.236	0.247	0.252	0.252
K_2ReO_4	3758	0.167	0.207	0.149	0.158	0.160	0.147	0.163	0.160	0.181	0.202	0.140	0.165	0.150	0.183
Iodine	3754	0.251	0.200	0.260	0.302	0.336	0.224	0.221	0.258	0.304	0.372	0.362	0.376	0.350	0.342
K_2PT(trimenthyl-dibenzyl amine)	3274	0.190	0.298	0.154	0.189	0.181	0.195	0.200	0.230	0.288	0.310	0.270	0.210	0.224	0.230
$PtBr_2(NH_3)_2$ + K_2ReO_4	3754	0.186	0.257	0.159	0.166	0.158	0.163	0.180	0.180	0.189	0.195	0.206	0.224	0.201	0.202
$PtBr_2(NH_3)_2$ + K_2Pt(trimethyl-dibenzyl amine)	3765	0.193	0.321	0.158	0.173	0.155	0.158	0.155	0.163	0.178	0.191	0.195	0.226	0.237	0.235

TABLE II

COORDINATES AND SUBSTITUTION PARAMETERS FOR GENE 5 DNA UNWINDING PROTEIN DERIVATIVES

Compound	X	Y	Z	A	B	Res.	$\frac{\Delta F''}{F + \Delta F'}$
$Pt(NH_3)_2Br_2$	0.371	0.0000	0.0917	61	19	2.4 Å	0.0614
Iodine	0.2270	0.2654	0.7629	34	6	2.3 Å	—
	0.5000	0.0549	0.0000	12	20		
$Pt(NH_3)_2Br_2$ + K_2Pt(trimethyl-dibenzylamine)	0.0357	−0.0047	0.0960	67	27	2.3 Å	0.0554
	0.0062	−0.0741	0.1040	22	20		
	0.2576	0.4002	−0.0362	6	30		
K_2ReO_4	0.5000	0.0462	0.0000	42	40	2.6 Å	0.0763
	0.4204	0.2983	−0.0454	9	10		
	0.0518	0.2465	0.0523	5	13		
	0.4025	0.0215	−0.2072	5	10		
K_2Pt(trimethyl-dibenzylamine)	0.0288	−0.0128	0.0970	40	45	2.6 Å	0.0443
	0.5000	0.1293	0.0000	9	36		
	0.5000	0.2566	0.0000	7	30		
	0.3854	0.4156	0.7099	4	5		
$Pt(NH_3)_2Br_2$ + K_2ReO_4	0.0344	−0.0055	0.1030	22	18	2.3 Å	0.0575
	0.5000	0.0407	0.0000	36	34		
	0.3027	0.9519	0.2053	3	20		

Figure of Merit to 2.3 Å = .73
Total Number of Reflections to 2.3 Å = 3793

squares minimization of the lack of closure. The program employed was that of Rossmann et al. (16) using the procedure of Dickerson et al. (17). The quality and phasing contribution of each derivative was evaluated by consideration of the residuals and statistics shown in Tables I and II. The election density map had a mean figure of merit of 0.72 and a distribution with sin θ/λ shown in Table III. Most parts of the map were clearly interpretable in terms of a continuous polypeptide chain although the N and C terminal residues and one loop showed obvious indications of disorder.

A Kendrew model of the gene 5 monomer was constructed using a Richard's optical comparator on a scale of 2 cm/Å from the 2.3 Å electron density map. Coordinates were measured for all nonhydrogen atoms with a plumb and line. The fitting of amino acid side-chains was in good agreement with the known sequence although a number of hydrophilic residues on the surface of the molecule were somewhat disordered or absent. We are at present in the process of refining the structure by least squares and difference Fourier technique which is not yet complete. Thus all of our conclusions at this time are based on the model constructed directly from the MIR electron density map.

TABLE III

FIGURE OF MERIT AND RESIDUAL DISTRIBUTION AS A FUNCTION OF RESOLUTION

Zone No.	Overall	1	2	3	4	5	6	7	8	9	10	11
Zone spacing A		7.71	5.35	4.74	4.23	3.63	3.32	3.11	2.88	2.69	2.51	2.38
R Modulus	0.605	0.614	0.517	0.547	0.589	0.604	0.675	0.642	0.588	0.569	0.607	0.765
R Weighted	0.473	0.457	0.328	0.365	0.423	0.428	0.542	0.516	0.499	0.474	0.614	0.933
Fig Merit	0.724	0.869	0.899	0.866	0.842	0.853	0.772	0.757	0.694	0.613	0.595	0.533
No. of F's	3481.	251.	147.	131.	245.	426.	265.	292.	491.	357.	578.	298.

Fig. 1 depicts a wooden model of the gene 5 protein at an effective resolution of ~5.0 Å viewed approximately down the crystallographic 100 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 Fig. 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 ~30 Å in length, run antiparallel courses and are separated by ~25 Å.

The course of the polypeptide chain in the gene 5 monomer is shown in Fig. 2 and 3 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 (16) and sequence-structure rules (17). There are three basic elements of secondary structure 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 loops (50–70) that creates the appendage of density near the molecular dyad and maintains the dimer species in solution. The second β-loop (71–82) forms the top surface of the molecule and we believe 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

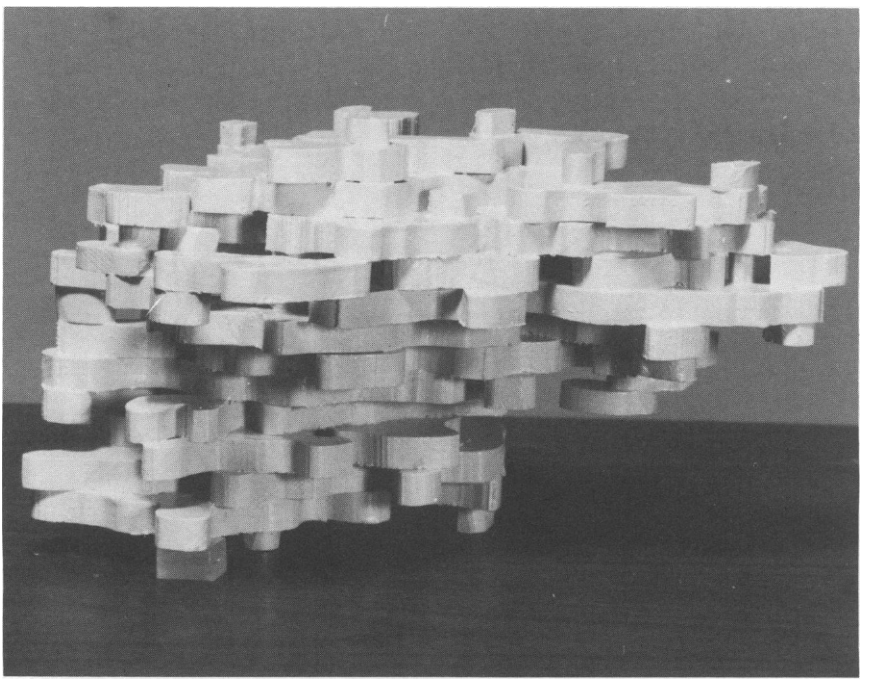

Figure 1 Representation of the gene 5 protein electron density based on the 2.3 Å Fourier made by cutting appropriate envelopes of density from each section of map and assembling them in the y direction. Model is viewed approximately along the 100 direction and can be seen as an essentially globular mass with a protrusion of density near the molecular dyad.

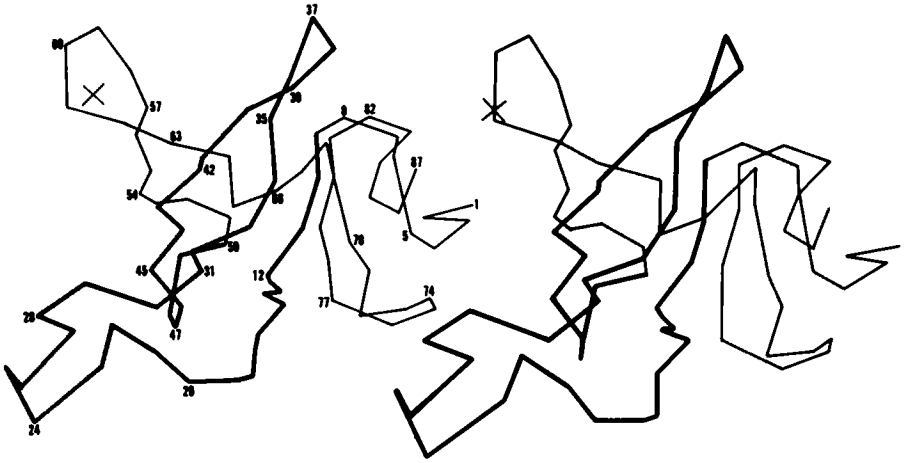

Figure 2 Stereo representation of the polypeptide backbone of the gene 5 protein using only β-carbon coordinates. The emphasized portion of the chain constitutes the three strands of the antiparallel β-pleated sheet primarily responsible for the binding of the DNA.

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.

The long groove beneath the three stranded sheet by its shape and extent suggests it to be the DNA binding interface. There is no other passage through the density that would be consistent with a long polynucleotide binding region. Given this to be the site, then the mode of cross strand attachment of the gene 5 protein would be that shown in Fig. 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 ~25 Å.

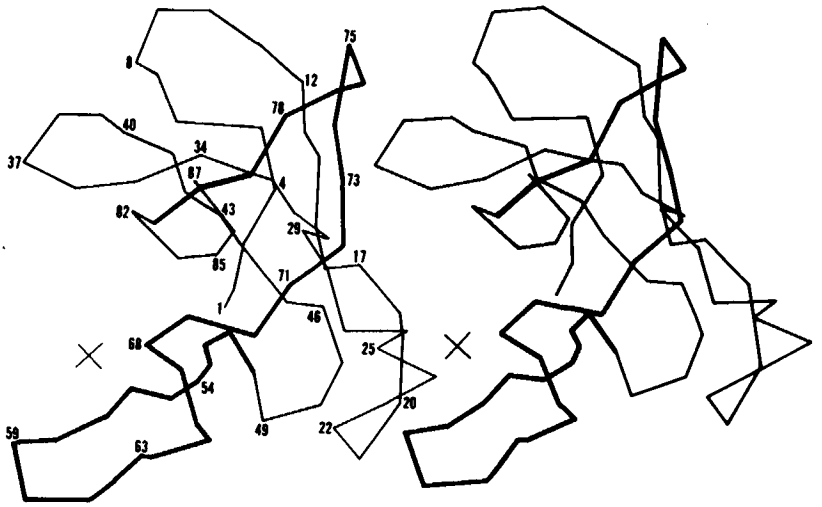

Figure 3 Stereo representation of the polypeptide backbone of the gene 5 protein rotated so that the view is roughly along the course of the DNA binding groove. This groove is ~25 Å in length and runs more or less parallel with the strands of the β-sheet.

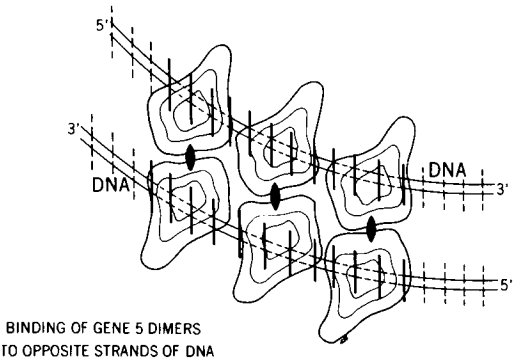

Figure 4 Schematic representation illustrating the cross-chain binding of the gene 5 dimers to opposing strands of a DNA duplex or opposite sides of a circular single-stranded DNA molecule. The distance between opposing DNA single strands would be ~25 Å.

THE DNA BINDING SITE

The binding cleft in the gene 5 protein is composed primarily of the amino acid side-chains arising from residues 12–49 of the antiparallel β-sheet shown in Fig. 5. These strands run more or less parallel with the direction of the DNA chain 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 space-filling drawing of the gene 5 monomer showing the binding region is shown in Fig. 6.

Aromatic amino acid side-chains have been inplicated in the binding of DNA to the gene 5 molecule by chemical modification and nuclear magnetic resonance (NMR) studies. These show that tyrosines 25, 41, and 56 lie near the surface of the protein and are readily substituted by tetranitromethane which prevents DNA binding (17). Conversely, binding of oligonucleotides or DNA before reaction prevents nitration of these residues. ^{19}F-NMR of the fluorotyrosyl containing protein confirms these results and further suggest that these tyrosines intercalate or stack with the bases of the DNA (18). Similar kinds of results have been obtained with deuterated protein that implicates at least one phenylalanine residue in a

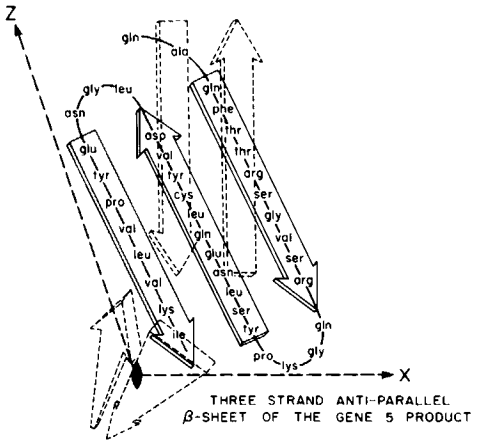

Figure 5 A schematic diagram showing the three components of β-structure that comprise the gene 5 protein. The amino acid residues forming the three stranded sheet are indicated. These amino acids are primarily engaged in interacting with the single stranded DNA chain.

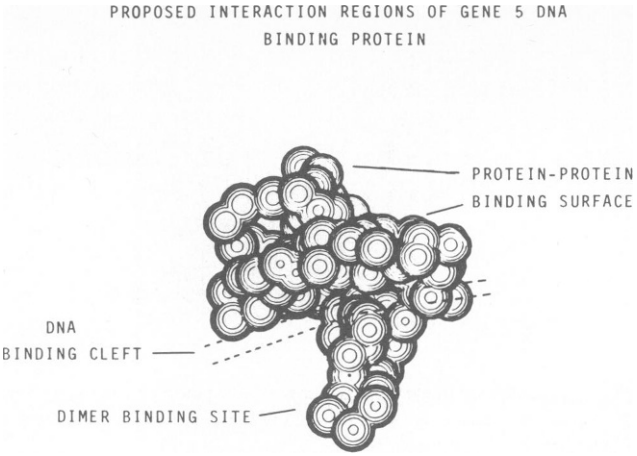

Figure 6 Drawing of the polypeptide backbone of the gene 5 protein viewed along the DNA binding groove with each β-carbon represented by a sphere of 3.0 Å diameter to give a space-filling effect.

similar fashion (19). Spectral data lend further support to the contention that aromatic residues of the protein stack upon or intercalate between bases of the DNA (20).

A number of aromatic residues are arrayed along the binding surface, and these include tyrosines 26, 41, 34, 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 nearest to the viewer in Fig. 6, contains only tyrosine-26. The aromatic side-chains, with the exception of tyrosine 56 and phenylalnine-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 tyrosines 41 and 34 and phenylalanine 13 which form a triple stack with Phe-13 most interior, Tyr-41 fully on the outside, and Tyr-34 interposed between. The stacking is also not precisely one atop the other, but the rings are fanned out like three playing cards. These side groups 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. (18) note from their NMR data that in the uncomplexed protein a number of tyrosyl proton resonances demonstrate 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 implying a disruption of the pattern as the residues begin interacting with the bases of the DNA.

Tyrosine-26 is near the turn between strands 1 and 2 of the antiparallel β-sheet. This bend appears to be a very flexible elbow of density extending out away 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 or one removed from a proline residue. The backbone structure of the protein is engaged in β-structure and one might expect that this hydrogen bonding network would restrict the freedom of bulky side groups. By virtue of their proximity to a natural structure disrupting amino acid, proline, however, these three tyrosines are endowed with more liberty than they

might otherwise possess. Because of the proline residues, the tyrosine side-chains can rotate from one side of the sheet to the other through a 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 (17). This is consistent with its location in the binding groove as is the finding that this -SH group can be photo crosslinked to thymidine residues of bound nucleic acid (21).

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 (17). 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 nucleic acid but do not form highly rigid salt bridges or hydrogen bonds (18). Resonances from the 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 (18).

The DNA binding trough has in its interior a rather large number of basic amino acid residues which, because of the length and flexibility of these side chains, reach into the groove though originating at disparate locations throughout the molecule. The basic residues most clearly apparent in the cleft are arginines 21, 80, 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 positively 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 is certainly in close proximity to the interface, but we see it turned away from the groove rather than toward it.

The binding cleft is very interesting in that the positively charged residues of lysine and arginine are distributed predominently over the most interior surface while the aromatic residues are arrayed primarily along the exterior edges. Thus is 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 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 (16) from micrograph and spectral data that the DNA in the gene 5 complex is completely unstacked and stretched along the filament axis and the demonstration that the adenine bases of DNA bound to gene 5 protein can be modified at their amino group (22). That small, but not gross, conformation changes occur in the protein upon DNA binding is in agreement with the NMR studies of Coleman et al. (18) on the α-CH and oliphatic methyl groups which suggest that gene 5 must contain a large percentage of fixed structure without large regions of flexible polypeptide chain. Day's (16) spectral evidence 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 (17). 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, is consistent with the physical and chemical studies. Therefore, although our binding mechanism is speculative, it is to our knowledge consistent with the structure as we visualize it, and the evidence at hand from noncrystallographic analyses.

CRYSTALLINE GENE 5 PROTEIN-DEOXYOLIGONUCLEOTIDE COMPLEXES

At least ten different crystal habits of gene 5 protein complexed with deoxyoligonucleotides have been observed in our crystallization trials. The dominant forms are rhombic plates, though triangular and hexagonal prisms and plates and crystals such as seen in Fig. 7 are also frequently encountered. We have commonly observed polymorphism in single samples and transformations between different crystal forms as well.

The oligonucleotides used in the crystallization experiments were d-pGpC, d-pApT, d-$(Ap)_4$ and d-$(Ap)_8$ from Collaborative Research, Waltham, Mass. The specific sequence oligomers d-pCpTpTpC and d-$(Tp)_4$ were gifts of Doctors Robert Ratliff and Lloyd Williams of the University of California Los Alamos Scientific Laboratory; the homopolymers d-$(Cp)_3$ and d-$(Cp)_4$ were gifts of Dr. Gobind Khorana of M.I.T. d-GGTAAT and its complimentary hexamer were supplied by Dr. Jack Van Boom of the University of Leiden.

The complexes were crystallized by the vapor diffusion method in depression plates again using polyethylene glycol, except that we found the complexes to grow more readily from PEG 6000 than PEG 4000. The samples were again buffered at pH 7.5 by 0.01 M Tris-HCl and

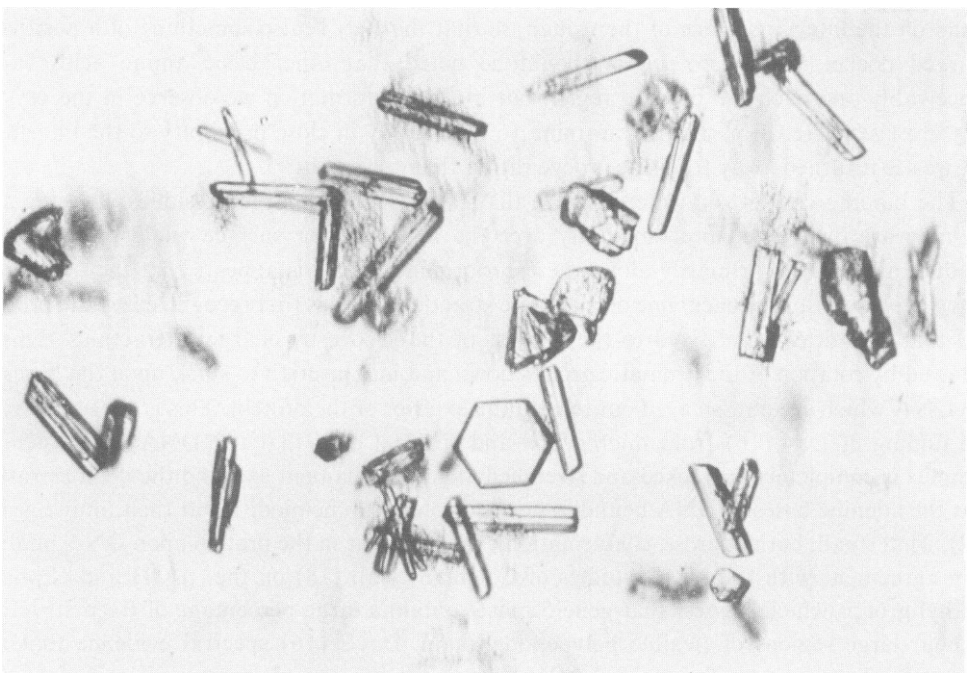

Figure 7 Low power light-microscope photograph of a lath modification of crystals formed from complexes between the gene 5 protein and the oligonucleotide d-(pCpTpTpC). A number of other habits have been observed of this same complex.

the final concentrations of PEG were in the range of 10–14%. Times for growth varied from 12 h in some cases to >3 mo in others.

The unit cell parameters and symmetry properties of four independent crystal modifications of the gene 5 protein-DNA complex are shown in Table IV. We noted that three of the crystals are based on hexagonal systems characterized by sixfold symmetry and the fourth, of space group $C222_1$, can be related to the $P6_3$ unit cell if one assumes a pseudo hexagonal packing arrangement. In fact, we frequently observe this orthorhombic crystal form growing as a twin or satellite with a crystal of hexagonal habit.

Although we could not measure the density of any of the complex crystals directly, we assumed a volume to mass ratio for each that was near the center of the range of crystalline proteins compiled by Matthews (23) and was consistent as well with that measured for the uncomplexed gene 5 protein crystals, V_m = 2.45 Å/dalton (13). Given this we estimate that the most reasonable number of gene 5 monomers in each asymmetric unit was consistently 12 (or 6 dimers), except for the $P3_1$ form in which we judged there to be ~24.

The crystals of space group $P3_1$ were the best crystals we examined and x-ray diffraction data to 5.0 Å resolution were collected from this form. The volume of the asymmetric unit of this crystal was twice that of the others, but the diffraction pattern of these trigonal crystals shows very high 32 pseudo symmetry and this suggested the presence of a near crystallographic twofold axis along the 100 or 110 directions in the crystal. Thus the effective asymmetric unit contains 12 gene 5 monomers, the same number determined for the other crystal forms.

In the $P6_3$, $C222_1$, R32, and pseudo $P3_121$ crystals the number of gene 5 monomers per asymmetric unit is observed to be ~12. The repeated occurrence of this number of monomers as the asymmetric unit of the crystals suggests rather strongly a specific aggregate of 12 gene 5 monomers that is formed upon addition of oligonucleotides to the protein. The fact that these aggregates crystallize requires that they be a homogeneous population of identically structured complexes and they must represent some ordered mode of self assembly from the solution species.

Three dimensional x-ray diffraction data were collected to 5.0 Å resolution on the trigonal crystals of the gene 5-oligonucleotide complexes again using the Picker FACS-I diffractometer. Only single measurements of each independent reflection were made using the step scan mode allowing, however, the entire data set of 8,000 reflections to be collected from a single crystal. To this data we applied the rotation and translation function in an attempt to

TABLE IV
CRYSTAL FORMS OF FD PHAGE GENE 5 PROTEIN COMPLEXED WITH OLIGODEOXYNUCLEOTIDES

Hexagonal plates	Diamond plates
a = 107	a = 110
	b = 180
c = 206	c = 117
$P6_3$	$C222_1$
12 * 9,800 daltons per asymmetric unit	12 * 9,800 daltons per asymmetric unit
Rhombohedra	Hexagonal prisms
a = 140	a = 143
	c = 83
α = 60	
R32	$P3_1$
12–18 * 9,800 daltons per asymmetric unit	24 * 9,800 daltons per asymmetric unit

determine the symmetry and the orientation of the gene 5 monomers in the asymmetric units of the complex crystals.

All computing operations were performed on a PDP11/40 computer. The rotation function was that of Crowther (24) as modified by Tanaka (25) for a spherical polar coordinate system. The translation function and structure factor calculation programs used were those of Lattmann (26). For rotation function calculations involving self vector searches within the native set, data between 10 and 6 Å resolution and with intensity >4 SD were employed. The maximum length of the vectors included in the search varied from 20 to 35 Å. For the searches of the complex using the native gene 5 protein structure as the search model, we used in succession only α-carbon atoms, all main chain atoms, and all main chain atoms plus tyrosines, phenylanines, methionines, and cysteine. Data between 10 and 6 Å resolution were used and the maximum vector length was varied between 20 and 35 Å.

The rotation function searches of the complex crystals have so far given rather inconclusive, though encouraging results. The searches using the self vectors and those from the native structure have revealed three local symmetry axes. The first of these, and most prominent, is a dyad axis nearly parallel to the crystallographic a* axis. This confirms our conclusion, based on the pseudo symmetry seen in the diffraction pattern, that the true space group of $P3_1$ contains molecules packed nearly, but not exactly, with the symmetry of space group $P3_121$ and that the pseudo-asymmetric unit contains 12 gene 5 monomers.

A second rotation function peak, indicative of a local sixfold peak also occurs in the search map, but at a point that might be explained by the interaction of the local dyad found above and the crystallographic threefold axis. We cannot, however, rule out the possibility that this peak arises from symmetry internal to the 12 monomer aggregate. The last major peak in the search map is indicative of a second local dyad axis lying approximately in the *ab* plane which has no other obvious explanation than that it is internal to the duodecamer.

There is evidence from solution studies that aggregation of gene 5 protein does occur in the presence of oligonucleotides as well as DNA. Rasched and Pohl (11) have found from suberimidate crosslinking and SDS gel electrophoresis of gene 5 protein combined with short oligonucleotides that polymeric protein species up to "about eight" are formed. The lack of certainty in their upper limit is due, at least in part, to the anomalous electrophoretic mobility of crosslinked protein aggregates which would be expected to undergo more rapid migration since they are not completely extended polypeptide chains. Thus the size of these aggregates is not inconsistent with the 12 monomer aggregate found in our asymmetric unit. In addition, the complex between gene 5 protein and fd phage DNA formed in solution and studied by electron microscopy shows "a helical rodlike structure" in which there are 12 gene 5 monomers per turn of the helix (1). We believe we are observing crystallographically a structure similar to that existing under physiological conditions.

The gene 5 protein binds to DNA in a linear and highly cooperative manner, i.e., successive gene 5 molecules tend to bind immediately adjacent to one already bound rather than to an isolated site. This apparently reflects the existence of strong protein-protein interactions between adjacent gene 5 molecules along the DNA strands, and may explain the powerful helix destabilizing effect exerted by the protein. These strong contiguous interactions do not occur between gene 5 molecules in solution in the absence of nucleic acid; if they did they would lead to aggregate formation of free molecules and this is not observed. It appears that the potential for forming such interactions is a consequence of conformational changes in the protein molecules induced by binding to DNA. The triggering of conformation change caused by DNA or oligonucleotide binding, the resulting cooperative interaction between protein

molecules, and concomitant aggregation of the protein is likely responsible for the asymmetric unit of 12 monomers that we observe in our complex crystals.

Virtually all protein oligomers and large protein complexes studied so far by x-ray diffraction analysis have demonstrated symmetry relationships, or at least a high degree of quasisymmetry, between the units involved. This seems likely to be the case with the gene 5 protein-oligonucleotide complex as well. We know from crystallographic studies on the free protein that the gene 5 dimers contain perfect dyad axes relating monomers in pairs. The occurrence of six of these dimers in the asymmetric unit of the crystals suggests the likelihood of an aggregate having sixfold symmetry. This is reinforced by the finding that three of the four unit cells encountered are of hexagonal symmetry and that the fourth can be interpreted in terms of hexagonal packing. Although there is no required correlation, objects with hexagonal symmetry do tend to express such symmetry in the crystalline state and the number of hexagonal forms observed in this case argues for such a correlation.

The aggregate occupying the asymmetric unit of the complex crystals, which in all unit cells so far examined has contained 12 gene 5 protein monomers, is a closed arrangement of fixed and determinate size which forms spontaneously in solution only when triggered by the binding of nucelic acid fragments. The simplest model for the asymmetric unit is that of a closed circle or disk having a sixfold axis along its center which is perpendicular to the twofold axes of the dimer units, i.e., it possesses 622 point group symmetry.

A MODEL FOR THE GENE 5-DNA COMPLEX

The shape of the gene 5 protein dimer in the unliganded state is known from x-ray diffraction analysis. The structure created when one takes these dimers and arranges them in a circle

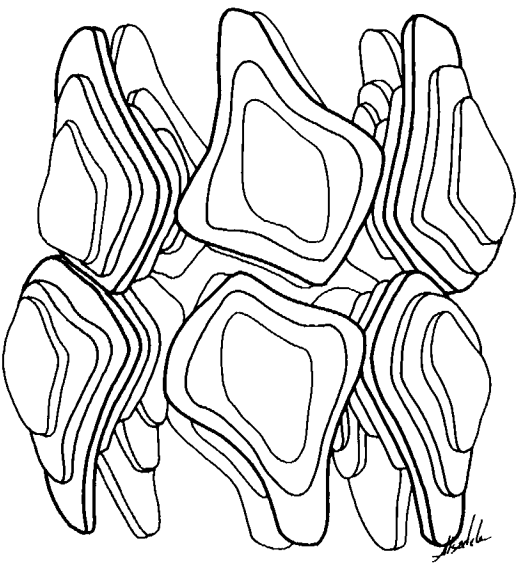

Figure 8 A proposed model for the asymmetric unit common to the four crystal forms of gene 5 protein DNA complexes we have analyzed. The arrangement is a circle, or disk, having 622-point group symmetry formed by joining the two ends of a linear array of six gene 5 dimers, each of which possesses an inherent dyad, to produce closure. The upper hexagon of monomers will bind a single strand of DNA running in one direction, and the lower hexagon of monomers will bind a strand running in the opposite direction. The disk has a diameter of ~100 Å and a height of ~80 Å. We believe that the DNA-binding region of each monomer faces the outside of the circle.

such that the dyad axes are perpendicular to a central sixfold axis is schematically shown in Fig. 8. Because of the double wing character of the gene 5 dimer, the aggregate would have a twofold crown shape with a diameter of ~100 Å and a thickness of ~80 Å. Thus it is not a flat disk shape but a squat cylinder. The aggregate we are proposing can be packed without difficulty in each of the unit cells we have characterized.

The aggregation phenomenon we seem to be observing with the gene 5 protein-oligonucleotide complexes is not unprecedented. The tobacco mosaic virus (TMV) disk (27) is an obvious analogy. Here again nucleic acid complexing proteins are stimulated upon nucleotide binding to form a helical rod. Unlike the aggregation seen with the fd gene 5 protein, aggregation of the TMV coat protein can be induced in the absence of nucleic acid by careful selection of the environment. Under these conditions the TMV protein molecules organize into a closed circle or disk having a 17-fold symmetry axis (28).

The aggregate of 12 gene 5 monomers observed in the crystal is not identical in structure to the helical aggregates of the gene 5 protein and DNA observed in the electron microscope. The unit we postulate in Fig. 5 is completely closed and does not allow an extended helix to be built up simply by translation along the direction of the sixfold axis. However, the relationship between the two structures may be somewhat analogous to the relationship that exists between the 17-fold TMV closed disk structure and the TMV helix which has an approximate

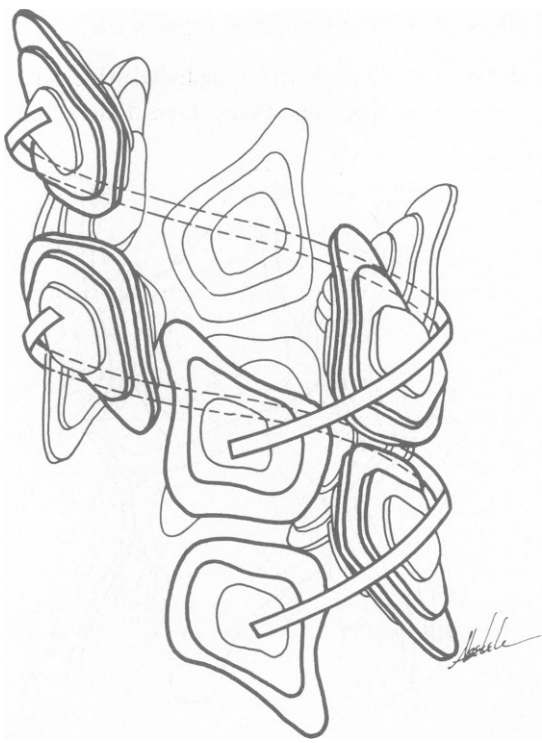

Figure 9 A proposed model for the structure of one turn of the gene 5 protein DNA double helix. This arrangement is produced by opening the disk structure (Fig. 9) between any two adjacent dimers and displacing the free ends along the unique axis direction. The stacking of these lockwasher units results in a double-helical structure that has a sixfold screw axis with perpendicular dyads, 12 gene 5 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 5 protein.

17-fold screw axis. The latter structure arises from the first simply by opening the disk and displacing the two ends along the direction of the unique axis to produce a "lock washer" unit. The free ends of these "lock washer" units are joined as the units are stacked to produce the helix. A model of a helical structure that might be produced by the gene 5 protein binding to two strands of DNA running in opposing directions is shown in Fig. 9. This helical structure would contain essentially the same lateral interactions between adjacent protein monomers as occur in the closed disk. This structure is a gene 5 double helix, one chain of which binds a DNA strand running 3′ to 5′ and the other a strand running 5′ to 3′. It has a sixfold screw axis with a linear repeat of ~80–90 Å and a diameter of ~100 Å.

This research was supported by grants from the National Institutes of Health, the National Science Foundation and the American Cancer Society. We thank Doctors J. H. van Boom, Lloyd Williams, and George Ratliff for supplying deoxyoligonucleotides.

Received for publication 3 December 1979 and in revised form 4 February 1980.

REFERENCES

1. Nakashima, Y., A. K. Dunker, D. A. Marvin, and W. Konigsberg. 1974. The amino acid sequence of a DNA binding protein, the gene 5 product of fd filamentous bacteriophage. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **40**:290.
2. Alberts, B., L. Frey, and H. Delius. 1972. Isolation and characterization of gene 5 protein of filamentous bacterial viruses. *J. Mol. Biol.* **68**: 139–152.
3. Oey, J. L., and R. Knippers. 1972. Properties of the isolated gene 5 protein of bacteriophage fd. *J. Mol. Biol.* **68**:125–128.
4. Mazur, B. J., and P. Model. 1973. Regulation of coliphage f1 single-stranded DNA synthesis by a DNA-binding protein. *J. Mol. Biol.* **78**:285–300.
5. Salstrom, J. S., and D. Pratt. 1971. Role of coliphage M13 gene 5 in single-stranded DNA production. *J. Mol. Biol.* **61**:489–501.
6. Cavalieri, S., D. A. Goldthwait, and K. E. Neet. 1976. The isolation of a dimer of gene 8 protein of bacteriophage fd. *J. Mol. Biol.* **102**:713.
7. Dunker, A. K., and E. A. Anderson. 1975. The binding of the fd gene-5 protein to single-stranded nucleic acid. *Biochim. Biophys. Acta.* **402**: 31–34.
8. Henry, T. J., and D. Pratt.1969. The proteins of bacteriophage M13. *Proc. Natl. Acad. Sci. U.S.A.* **62**:800.
9. Pratt, D., P. Laws, and J. Griffith. 1974. Complex of bacteriophage M13 single-stranded DNA and gene 5 protein. *J. Mol. Biol.* **82**:425–439.
10. Gray, C. W., R. S. Brown, and D. A. Marvin. 1979. Direct visualization of adsorption protein of fd phage. European Molecular Biology Laboratory. ICN-UCLA Symposium on Recognition and Assembly in Biological Systems, Keystone, Colo.
11. Pretorius, H. T., M. Klein, and L. A. Day. 1975. Gene V protein of fd bacteriophage. Dimer formation and the role of tyrosyl groups in DNA binding. *J. Biol. Chem.* **250**:9262.
12. Rasched, I., and F. M. Pohl. 1974. Oligonucleotides and the quarternary structure of gene-protein from filamentous bacteriophage. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **46**:115.
13. McPherson, A., I. Molineux, and A. Rich. 1976. Crystallization of a DNA-unwinding protein: Preliminary x-ray analysis of fd bacteriophage gene 5 product. *J. Mol. Biol.* **106**:1077.
14. McPherson, A., F. A. Jurnak, A. H. J. Wang, I. Molineux, and A. Rich. 1979. Structure at 2.3 Å Resolution of the Gene 5 Product of Bacteriophage fd: a DNA unwinding protein. *J. Mol. Biol.* **134**:379–400.
15. Wyckoff, H. W., M. Doscher, D. Tpernoglon, T. Inagami, L. N. Johnson, K. D. Hardmann, N. M. Allewell, D. M. Kelly, and F. M. Richards. 1967. Design of a diffractometer and flow cell system for x-ray analysis of crystallized proteins with applications to the crystal chemistry of RNase-S. *J. Mol. Biol.* **27**:563.
16. Adams, M. J., D. J. Haas, B. A. Jeffrey, A. McPherson, H. L. Mermal, M. G. Rossmann, R.W. Schevitz, and A. J. Wonacott. 1969. Low resoltuion study of crystalline L-lactate dehydrogenase. *J. Mol. Biol.* **41**:159.
17. Dickerson, R. E., J. C. Kendrew, and B. E. Straudberg. 1961. The crystal structure of myoglobin: phase determination to a resolution of a 2 Å by the method of isomorphous replacement. *Acta Cryst.* **14**:1188.
18. Day, L. A. 1973. Circular dichroism and ultraviolet absorption of a deoxyribonucleic acid binding protein of filamentous bacteriophage. *Biochemistry.* **12**:5329.
19. Anderson, R. A., Y. Nakashima, and J. E. Coleman. 1975. Chemical modifications of functional residues of fd gene 5 DNA-binding protein. *Biochemistry.* **14**:907.

20. Coleman, J. E., R. A. Anderson, R. G. Ratcliffe, and I. M. Armitage. 1976. Structure of gene 5 protein-oligodeoxynucleotide complexes as determined by ^{1}H, ^{19}F, and ^{31}P nuclear magnetic resonance. *Biochemistry.* **15**:5419.
21. Coleman, J. E., and I. M. Armitage. 1979. Tyrosyl-base-phenylalanyl intercalation in gene 5 protein-DNA complexes: ^{1}H NMR of selectively deuterated gene 5 protein. *Biochemistry.* In press.
22. Nakashima, Y., and W. Konigsberg. 1975. Reinvestigation of a region of the fd bacteriophage coat protein sequence. *Int. Symp. Photobiol.*
23. Kohwi-Schigematsu, T., T. Enomoto, M. Yamada, M. Nakanishi, and M. Tsuboi. 1978. Exposure of DNA bases induced by the Interaction of DNA and Calf thymus DNA-helix-destabilizing protein. *Proc. Natl. Acad. Sci. U.S.A.* **75**:4689–4693.
24. Matthews, B. W. 1968. Solvent Content of Protein Crystals. *J. Mol. Biol.* **33**:1491.
25. Crowther, R. A. 1972. Fast rotation function. *In* The Molecular Replacement Method. M. G. Rossman, editor. Gordon and Breach, New York. 173–178.
26. Tanaka, H. 1977. Representation of the fast-rotation function in a polar coordinate system. *Acta Cryst. A.* **33**:191–193.
27. Lattman, E. E. 1969. Application of Molecular Replacement Methods to the Structure of Hemoglobin. Ph. D. Thesis. Johns Hopkins University, Baltimore, Maryland.
28. Butler, P. J. G. 1971. The mechanism and control of the assembly of TMV from its RNA and protein disks. *Cold Spring Harbor Symp. Quant. Biol.* **36**:461–468.
29. Champness, J. N., A. C. Bloomer, G. Bricogne, P. J. G. Butler, and A. Klug. 1976. The structure of the protein disk of tobacco mosaic virus to 5 Å resolution. *Nature (Lond.).* **259**:20.

DISCUSSION

Session Chairman: David Eisenberg *Scribe*: Pieter De Haseth

STUBBS: Protein nucleic acid binding on the gene 5 system and the TMV system have some interesting analogies. First, according to your paper the direct phosphate binding seems to be entirely by arginine, not lysine, and this is of course what we observe in TMV. I have never liked lysine as a candidate for binding nucleic acids. A few years ago Cosson et al. (1973, *J. Amer. Chem. Soc.* **95**:7) pointed out the hydrogen bonding capacity of arginine, which makes it a much better candidate for binding nucleic acids. However, although that part of our binding site seems to be similar, the base binding seems to be totally different. You get the base binding that everybody would expect, the bases lying flat against aromatic side chains, while we find aliphatic side chains instead.

Second, I rather like the fact that you want to bind the DNA backbone first, haul it into place, and then tie down the bases. This could be effectively the situation in TMV as well, since we observe from the work in Cambridge that as the RNA sits on top of the growing virus rod, the arginines from the incoming disk reach out to catch the phosphates. We are being as speculative as you are, but it seems highly likely that in both cases the electrostatic anchors are made first, and then the fine detail is put in by the bases.

MCPHERSON: I might add that we also find a considerable number of aliphatic residues along the binding crevice. In the interior of the crevice, however, are primarily the arginines and the lysines. One difficulty we have is that our structure is not yet highly refined. We are perhaps stretching things a bit when we say that there are arginines in close contact with the DNA, and we don't see many lysines. But the arginines at this level are obvious to us, at least, and we'd be hard pressed to say that they are not involved in phosphate binding. The lysines may be, but that is not as certain.

BUTLER: It is not obvious that the mechanism of nucleic acid binding is going to be the same during the initiation step of TMV assembly and during the subsequent elongation. The initiation is highly sequence-dependent and it may well be that there is a recognition. It is not even clear that it is exactly the same site on the protein that is involved in both. Present, very tentative, identification of the nucleotide bound into the disk in the crystal puts it under a different helix (of the protein) than the one Gerald Stubbs has identified in the virus.

On another point, I was very interested in your Fig. 4. We have found an approximate dyad axis in TMV coat protein, and we speculate that it may have evolved from a two-helix protein, which in a very primitive virus actually formed pairs on each side of the RNA; these stacked back-to-back to give the virus helix. I find it interesting that you may have one doing a similar thing.

BERGET: I was struck by Fig. 9 of your paper for a reason which stems more from the biology of the virus rather than from the elegant crystal structure. Until today my naive view of the gene 5 protein of fd was that it was simply a

single-stranded DNA-binding protein. In this figure, however, it is majestically elevated to a scaffolding protein for the formation of the rodlike final virus structure. I am also concerned whether this model takes into consideration the fact that there is a very mysterious displacement reaction that now has to occur. In this respect the gene 5 protein finds itself in the same trouble as the scaffolding protein in P22 virus, in that the fd gene 5 protein is now on the inside of the structure and has to leave as the coat protein is put around the outside of the structure. You might be able to comment briefly on that. I have a question that you might bring in: what does the final structure of the DNA in the virus look like compared to the structure of DNA in this coiled spool that you propose?

MCPHERSON: I cannot answer your last question, but there are other people here who could. Let me comment on your first point, that the gene 5 protein-DNA complex appears in Fig. 9 to be a double-stranded structure. This is absolutely true. The protein binds to single-stranded DNA, but to one side of the DNA double helix. It then pulls it apart, and the opposite monomer binds to the opposite strand. So, because of the dyad axis which relates the two monomers, the dimer can handle the strands running 3′–5′ and 5′–3′ on opposite sides of the duplex. The final complex will in fact be a double-stranded complex, using the two halves of the helix. In the case of the single-stranded fd phage it will use the two sides of the single circle and will pull that circle together, making a helical rod. This of course is what Alberts and Gray observed in the electron microscope.

You are right about the scaffolding problem, if you are talking about the problem of getting the gene 5 protein off and the gene 8 protein on. Our model does have the DNA on the outside of the complex, so that the gene 5 protein in fact forms a spindle around which the DNA spools, in much the same way as in visualized in nucleosomes. The DNA is left on the outside of that complex and is fully exposed, I think, to approaches by the gene 8 protein.

The gene 8 protein is only half the molecular weight of the gene 5 protein. It also forms a helical structure but it obviously binds in a very different way because it is only half the molecular weight, and entirely α-helical. It binds much more tenaciously than the gene 5 protein, so it could pull the DNA off of the gene 5 protein. Also, all of this reaction occurs in the host cell membrane or close to the membrane where there are going to be some unusual environmental effects on the complex. It is therefore reasonable to think that the complex of gene 5 protein and DNA will fall apart under the influence of the gene 8 protein, and that a second structure will be formed.

BERGET: I agree with everything you said. It should be emphasized that as the gene 8 protein grabs onto the helix, there is gene 5 protein on the inside of this very complex structure. Getting that protein, which is twice as big as the gene 8 protein, out of that structure is a problem similar to that faced by a lot of double-stranded DNA viruses. In T4 the protein is hydrolyzed down to small peptides, and in P22 it is magically expelled and recycled.

MAKOWSKI: I am not at all happy with the designation of the gene 5 protein as a scaffolding protein. In the sense that the term "scaffolding protein" has been used in assembly in the past, this is probably not a place to use it, because the gene 5 protein is twice as large as the coat protein. The gene 5 assembly, as pictured in the figures in this paper, is actually several times the diameter of the complete virus particle. The gene 5 protein comes off the DNA inside the bacterial membrane, so that it actually comes off before the gene 8 protein (which is inserted with one end on each side of the membrane) interacts with the DNA.

There are several differences between the gene 5 complex and the intact phage worth noting here. The gene 5 complex has a specific 4:1 ratio of nucleotides to gene 5 proteins, whereas in the complete fd virus particle there are 2,3 nucleotides per coat protein monomer. The symmetry of the DNA is different from the coat protein in the complete virus particle. As for the assembly process that takes place in the membrane, in order to get a good handle on that we need to produce an image of the structure of the fd virus and, if possible, of the structure of the fd coat protein in the membrane as well as the image of the gene 5 complex which you have shown here.

BOTT: As I read your paper, it seems that gene 5 protein has a relatively unique structure in that it has a charged pocket, which is relatively interior, and hydrophobic residues arranged on the surface of the groove, relatively exterior. In light of the work on forces already discussed at this meeting, I wonder how the conformation is stabilized.

MCPHERSON: If you look at the amino acid sequence of the protein, you see that it has a very high proportion of hydrophobic amino acids all over. A very small protein of the 10,000 mol wt is not going to have the opportunity to form a large hydrophobic interior core. This, in part, explains a lot of the aggregation properties observed for the gene 5 protein. There are certainly large hydrophobic surfaces on the exterior of this protein: not only those hydrophobic residues which border the crevice in which nucleotides are bound, but also the backside of the protein, the C-terminal's 20 amino acids, has a very large hydrophobic surface. So in this protein, without any question, there is a lot of hydrophobic surface.

DAY: I would like to reiterate what Lee Makowski said about the role of this particular complex as a scaffolding. In addition to the major coat protein, which is in the membrane prior to assembly, there have recently been discovered two additional proteins, bringing the total number of proteins in the virus to 4. There are a couple of other genes

involved too, so the actual assembly mechanism for such a simple virus is quite involved. As Lee Makowski mentioned, the assembly of the virus takes place at the membrane; and as Alex McPherson mentioned, the gene 5 protein is recycled. So the actual assembly of the virus can be envisioned as a gigantic subunit exchange reaction.

As for the DNA structure in the gene 5 protein complex, a question I would like to ask has to do with the additional function of this protein. As the progeny single-strands are synthesized they are bound by this protein and protected against complementary strand formation. This is a rolling circle-type DNA synthetic machinery. Part of the progeny single-strand synthesis is a closure into the completed ring. As the ends are held together during the synthesis there has to be a slippage of the first synthesized DNA. What would your bet be as to whether the backsides of the protein in the dimers are slipping past the DNA or—what seems more likely—that the DNA is snaking through the subunits?

MCPHERSON: I think it is very unlikely that the DNA slips through the binding groove; it is going to be pretty tightly bound in there. I think it is the backside interactions that you are looking for because we do see a large hydrophobic surface on the exterior of the protein. This is going to form the primary protein-protein contacts. I think that slippage occurs here.

A word of caution before the next question. The model structures in Figs. 8 and 9 are speculative, and not based on hard crystallographic evidence. We have not solved those structures, although we are in the process of doing that. We have collected 5 Å data on one of these complex crystal forms. We are applying molecular replacement techniques to the solution. The models presented are those most consistent with all the evidence we have seen to date, but they are speculative.

LIPPARD: Concerning the heavy atom derivative which is written as K_2ReO_4, are you using the ReO_4^{2-} ion or is that supposed to be K ReO_4? I have noticed that the MIT group has published the K_2ReO_4 formulation.

I'm curious about the stacking interaction that you seem to suggest ought to be there between the aromatic amino acids and the DNA structure. From what you can see in the structure of the gene 5 protein, there is going to be an ionic interaction. What is the reason to believe that a stacking or intercalative interaction should exist?

MCPHERSON: Loren Day is in a better position to answer that than I am, because he has made many measurements showing where these stacking interactions between the aromatic side groups of the protein and the DNA bases occurred. But my understanding is that there is a considerable amount of evidence from NMR studies by Coleman and Armitage at Yale, and also spectroscopic evidence, that there is an interaction between the aromatic side groups on the proteins and the bases of the DNA. We have not yet directly visualized these interactions. We would like to bind some dinucleotide or larger fragment to the protein and see those interactions. We very recently have been able to get a monomer, GMP, to bind to the protein in our native crystals, and are presently collecting data on that complex. I am not certain a simple monomer is going to tell us that much, however.

FLETTERICK: What can you tell us about the conformational state of the protein with and without DNA? In a sense it is a self assembly process again.

MCPHERSON: Clearly the protein does not aggregate in the absence of oligomers or DNA, even at concentrations of up to 20 mg/ml of protein. As soon as dinucleotdies or longer oligonucleotides are added, aggregation appears to take place. What we must postulate is that the protein responds to the approach of the DNA by changing its conformation. The conformational change must be triggered by the binding of the DNA. We do not get cooperative interactions prior to the binding.

KALLENBACH: Do you have a feel for the number of nucleotide residues, either in the GMP case or in oligonucleotides, that actually saturate the binding site?

MCPHERSON: No, I do not. We can trigger the formation of the dodecamer with something as small as a dinucleotide and as large as a hexamer. Either there are (in the latter case) loose ends sticking out on the side, or the hexamer can crosslink adjacent dimers. We are not able to make a good correlation between the lengths of the oligomers that we use, or their sequence, and the particular crystal form that we get.

WELLS: I would like to know about the DNA. If you are pulling it in electrostatically, I would think that you are changing the orientation between the bases, perhaps changing the sugar pucker and certainly changing the phosphodiester bonds.

MCPHERSON: I don't know about the sugar pucker, but certainly spectroscopic evidence shows that when the DNA is bound to the gene 5 protein the bases are totally unstacked. The DNA is in a fully extended conformation. One presumes that as it is tearing up the double helix the protein is also pulling apart the bases. In some cases the bases are still open to chemical modification. For example, one can modify the amino groups on the adenines of the DNA when it is bound to gene 5 protein.

DAY: I want to ask you about the sulfhydryl group on the gene 5 protein, which is always in the reduced state in the native protein. Do you see this at the present level of resolution and is it involved in the DNA binding site?

MCPHERSON: That is a very interesting residue. We have hoped that we could react the single sulfhydryl group with mercury derivatives, in order to speed up the elucidation of the structure of the protein. In fact, that sulfhydryl group has proven to be almost completely inaccessible to us. We have never been able to form a derivative. Yet we see the group quite clearly in the electron density map. It is inside the deepest part of the crevice that binds the DNA, but it is turned around facing the hydrophobic region that exists in the interior core of the protein. In solution one can get at the group with $Mg(Ac)_2$ or something small like that. This destroys the binding of the DNA completely, as expected from our map.

DAY: It is interesting that the phosphates are probably inside and the bases outside in the DNA groove. As to the arrangement of DNA in fd virus, we have gone through a line of reasoning based on deductions for possible DNA structures in Pfl and Xf virus and concluded that in fd virus the phosphates might be inside. The precursor structure (the gene 5:DNA complex) might serve to arrange the DNA in this general way, although here the DNA is around the protein whereas in the virus the protein surrounds the DNA. I want to add, though, that David Banner and Don Marvin in Heidelberg have very recently obtained diffraction patterns for fd virus which have spots assignable to DNA that they interpret in terms of a classical type DNA structure, with phosphates outside. Finally, I certainly would like to congratulate you on your lovely work on the gene 5 precursor complexes.