Title
Structure at 2.3 Å resolution of the gene 5 product of bacteriophage fd: A DNA unwinding protein

Journal
Journal of Molecular Biology, 134(3)

ISSN
0022-2836

Authors
McPherson, Alexander
Jurnak, Frances A
Wang, Andrew HJ
et al.

Publication Date
1979-11-01

DOI
10.1016/0022-2836(79)90359-0

J. Mol. Biol. (1979) **134**, 379–400

Structure at 2·3 Å Resolution of the Gene 5 Product of Bacteriophage fd: A DNA Unwinding Protein

ALEXANDER MCPHERSON[1], FRANCES A. JURNAK[2], ANDREW H. J. WANG[2]
IAN MOLINEUX[3] AND ALEXANDER RICH[2]

[1] *The Department of Biological Chemistry*
The Milton S. Hershey Medical Center
The Pennsylvania State University
Hershey, PA 17033, U.S.A.

[2]*The Department of Biology*
Massachusetts Institute of Technology
Cambridge, Mass., U.S.A.

[3]*Imperial Cancer Research Fund*
Mill Hill Laboratories
London, England

(Received 10 October 1978, and in revised form 15 May 1979)

The structure of the gene 5 DNA unwinding protein from bacteriophage fd has been determined by X-ray diffraction analysis of single crystals to 2·3 Å resolution using six isomorphous heavy-atom derivatives. The essentially globular monomer appears to consist of three secondary structural elements, a radically twisted three-stranded antiparallel β sheet and two distinct antiparallel β loops, which are joined by short segments of extended polypeptide chain. The molecule contains no α-helix. A long groove, or arch, 30 Å in length is formed by the underside of the twisted β sheet and one of the two β ribbons. We believe this groove to be the DNA binding region, and this is supported by the assignment of residues on its surface implicated in binding by solution studies. These residues include several aromatic amino acids which may intercalate or stack upon the bases of the DNA. Two monomers are maintained as a dimer by the very close interaction of symmetry related β ribbons about the molecular dyad. About six residues at the amino and carboxyl terminus are in extended conformation and both seem to exhibit some degree of disorder. The amino-terminal methionine is the locus for binding the platinum heavy-atom derivatives and tyrosine 26 for attachment of the major iodine substituent.

1. Introduction

The gene 5 product of bacteriophage fd is a small DNA binding protein of 9800 molecular weight containing 87 amino acids (Alberts *et al.*, 1972; Oey & Knippers, 1972). Its primary physiological role is the stabilization and protection of single-strand DNA daughter virions from duplex formation following replication in the host (Salstrom & Pratt, 1971). Because of the highly co-operative nature of its binding to DNA, it further has the capacity to unwind or destabilize native DNA. This can be seen as a reduction by nearly 40 deg. C of the melting temperature for

0022-2836/79/310379-22 $02.00/0

double-stranded DNA in the presence of the gene 5 product (Salstrom & Pratt, 1971).

The protein is made in about 100,000 copies per infected *Escherichia coli* cell and can be isolated in substantial amounts by DNA cellulose chromatography. Its sequence (see Fig. 1) has been determined (Nakashima *et al.*, 1974) and extensive biochemical and biophysical characterization has been carried out (Coleman *et al.*, 1976; Anderson *et al.*, 1975; Day, 1973; Pretorius *et al.*, 1975). Evidence from these studies indicates electrostatic interactions between basic residues of the protein and the phosphate groups of the polynucleotide backbone are involved in binding, and that aromatic residues are likely to intercalate or stack upon the purine and pyrimidine bases during complex formation (Coleman *et al.*, 1976; Pretorius *et al.*, 1975).

Amino acid sequence of the gene 5 protein

MET-ILE-LYS-VAL-GLU-ILE-LYS-PRO-SER-GLN-ALA-GLN-PHE-THR-THR-ARG-SER-
10
GLY-VAL-SER-ARG-GLN-GLY-LYS-PRO-TYR-SER-LEU-ASN-GLU-GLN-LEU-CYS-TYR-
20 30
VAL-ASP-LEU-GLY-ASN-GLU-TYR-PRO-VAL-LEU-VAL-LYS-ILE-THR-LEU-ASP-GLU-
40 50
GLY-GLN-PRO-ALA-TYR-ALA-PRO-GLY-LEU-TYR-THR-VAL-HIS-LEU-SER-SER-PHE-
60
LYS-VAL-GLY-GLN-PHE-GLY-SER-LEU-MET-ILE-ASP-ARG-LEU-ARG-LEU-VAL-PRO-
70 80
ALA-LYS

FIG. 1. Amino acid sequence of the gene 5 protein (from Nakashima *et al.*, 1974).

Electron microscopy studies on the gene 5 protein (Alberts *et al.*, 1972; Pratt *et al.*, 1974), which exists predominantly as a dimeric species in solution (Cavalieri *et al.*, 1976), suggest that the protein binds to DNA strands running in opposite directions so that it would crosslink opposing strands of a duplex or opposite sides of circular single strands of DNA. The mechanism for DNA duplex unwinding is simply a linear aggregation along the two opposite strands deriving from the highly co-operative nature of the lateral binding interaction (Dunker & Anderson, 1975). This high degree of co-operativity presumably is a product of strong protein–protein forces between adjacent molecules of the gene 5 protein along the DNA strands.

Approximately two years ago we were fortunate in being able to obtain this protein in a form suitable for single-crystal X-ray diffraction analysis (McPherson *et al.*, 1976). Since these crystals marked the first time that a truly DNA interactive protein had presented itself for such analysis we began a three-dimensional study. This we hoped would allow us, both from the native structure and from crystals of gene 5–DNA complexes, which we have now obtained (unpublished results), to deduce with high precision the atomic interactions which mediate the non-specific recognition and binding of DNA by protein molecules.

The crystals are of monoclinic space group $C2$ with one gene 5 monomer per asymmetric unit, immediately implying that the gene 5 dimer contains a perfect molecular dyad axis relating its two halves. The unit cell dimensions are $a = 76{\cdot}5$, $b = 28{\cdot}0$, $c = 42{\cdot}5$ Å and $\beta = 108°$. The resolution of the diffraction data extends to at least 1·2 Å Bragg spacings, and radiation damage to the crystals appears negligible for up to 60 hours or more.

We report the structure determination of these crystals using the isomorphous replacement technique and describe some of the initially interesting features of the molecule. We will defer a detailed description of the orientations and interactions of the amino acid side-chains until some degree of refinement has been completed.

2. Methods

E. coli strain K12 were grown to a density of 4×10^8 cells/ml and were infected at a multiplicity of 20 with wild type fd bacteriophage. The cells were harvested 2 h after infection and lysed in a French press. The gene 5 protein was prepared according to the procedures of Alberts *et al.* (1972), the primary steps being denatured DNA-cellulose and DEAE-cellulose chromatography. The protein was judged homogeneous by sodium dodecyl sulphate/polyacrylamide gel electrophoresis according to Laemmli (1970). Only a single band was observed migrating at approximately 10,000 M_r with virtually nothing corresponding to the dimer species. This was true of crystals used in the diffraction analysis as well.

Crystallization was carried out using the vapor diffusion technique (McPherson, 1976*a*) on Corning glass spot plates enclosed in sealed plastic boxes. The sample volume was 20 μl and the reservoir contained 25 ml. Crystals were grown in 2 ways. Initially, the protein was totally precipitated by combining 10 μl of a 15 mg/ml protein solution containing 0·05 M-Tris·HCl at pH 7·6 with 10 μl of 20% polyethylene glycol (PEG) 4000 and leaving the samples in equilibrium with reservoirs of 10% PEG 4000 at 25°C. Later, crystals were grown from non-precipitated protein by combining 10 μl of a 15 mg/ml protein solution containing 0·01 M-Tris·HCl at pH 7·6 with 10 μl of a 10% PEG 4000 solution and allowing the samples to equilibrate with a 12% PEG 4000 reservoir at 4°C. The crystals grown from the 2 sets of conditions were isomorphous.

For both heavy-atom substitution analysis and data collection itself, crystals of the gene 5 protein were mounted by conventional means in quartz capillaries containing a small amount of mother liquor and sealed with dental wax. Isomorphous derivative crystals were obtained by diffusing a variety of heavy-atom-containing compounds into previously grown crystals with addition of a few microlitres of a concentrated solution directly to the mother liquor. The iodine derivative was produced by direct addition of 5 mM-KI and 5 mM-I_2 to crystals for a period of 5 days before data collection. The heavy-atom derivatives search was conducted using Buerger precession cameras with a crystal to film distance of 90 mm. The X-ray source was an Elliott rotating anode generator operated at 40 kV and 40 mA with a focal spot size of 200 μm^2. The X-radiation was nickel filtered CuK_α with 18 h exposures at 25°C.

The photographic trials were converted to sets of integrated intensities *via* an Optronics P-1000 high-speed rotating-drum microdensitometer using a 100-μm raster size and interfaced directly to a PDP11/40 computer. The program employed for on-line integration of the optical density measurements was SCAN 11 written by Paul Bethge at Harvard University and modified by ourselves. All other computing operations were performed by remote job entry to an IBM370/168 at Pennsylvania State University at University Park.

Before full 3-dimensional data collection was initiated on any derivative, it was deemed necessary only that a major zone zero level diffraction photograph demonstrate an average change in the structure amplitudes greater than 15% of the mean when scaled against an equivalent native photograph. This appears to have been a fair criteria as all but one of the presumptive derivatives when collected in 3 dimensions yielded either interpretable difference Patterson maps and/or difference Fourier maps and could be refined successfully.

Three-dimensional X-ray diffraction data were collected on a Picker FACS-1 diffractometer fitted with a 1600 W Phillips fine focus X-ray tube to 2·3 Å using the step scan mode (Wyckoff *et al.*, 1967). Friedel pairs were recorded at $\pm 2\theta$ for all compounds and, in general, the 3700 independent reflections could be obtained from 1 or at most 2 crystals. In addition, each data set was collected twice on different crystals and averaged with

merging residuals of no more than 3·5% on F. Errors were measured from counting statistics (Arndt & Willis, 1966). Scaling of derivative to native structure amplitudes was carried out in shells of $\sin^2\theta/\lambda^2$ and the R factors for each compound used in isomorphous replacement are shown in Table 1. The overall residuals are entirely reasonable and indicate good substitution. Except for iodine, the distributions do not suggest any appreciable non-isomorphism.

A number of problems were encountered in the location of the heavy-atom substitution sites arising from their close distribution about the 2-fold axis. The difference Patterson peak yielded by $PtBr_2(NH_3)_2$ seen in the Harker section of Fig. 2 was nearly 3 times the height of any other, but its proximity to the origin suggested caution. It was, however, confirmed by the anomalous difference Patterson of Fig. 3. A single isomorphous native Fourier based on the $Pt(NH_3)_2Br_2$ derivative was calculated but proved to be only marginally interpretable. The molecular envelope was clear and extensive regions of polypeptide chain could be traced, but we could by no means deduce the course of the chain.

Although the difference Patterson for K_2ReO_4 showed no solution, a difference Fourier based on the $PtBr_2(NH_3)_2$ derivative gave its position as a peak nearly 5 times above background at $x = 0{\cdot}50$, $y = 0{\cdot}08$, $z = 0{\cdot}00$. Its location on the dyad once again appeared suspicious; however, phases based on this site gave back $Pt(Br)_2(NH_3)_2$ in a difference Fourier. The validity of both the rhenium and platinum sites were confirmed by the difference Patterson between them (Rossmann, 1960) and is shown in Fig. 4. The major and minor iodine sites were readily found from difference Fourier maps as peaks 2 to 4 times background level and were confirmed by the corresponding difference Patterson syntheses. Additional difference Fourier rounds using all compounds and various combinations permitted location of the minor sites for all derivatives.

While the degree of substitution and the isomorphism of the derivatives appears to be quite good, the overall distribution of heavy-atom sites for the gene *5* crystals is rather poor in that most of the positions lie clustered very close to $x = 0{\cdot}00$ and $z = 0{\cdot}00$. Thus we do not expect that the quality of phasing by these derivatives is as good as might be expected if they were distributed at completely general sites. The iodine derivative shows obvious signs of non-isomorphism and was not used beyond 3·0 Å resolution.

The phase angles for the native structure were calculated using the isomorphous phasing method of Dickerson *et al.* (1961) and the heavy-atom parameters were refined with the error treatment of Blow & Crick (1959) to the values seen in Table 2. The program employed for the calculations was that written by Rossmann and his colleagues (Adams *et al.*, 1969). Phase angles were calculated at 5° intervals and cycles of phasing and least-squares minimization of lack of closure were alternated. The refinement was carried out in shells of $\sin\theta$ and the relevant statistics are shown in Table 3. Anomalous dispersion contributions were incorporated in the phasing for all derivatives except iodine as described by North (1965). The figure of merit and refinement residuals as a function of resolution are shown in Table 4 and the distribution of terms as a function of figure of merit in Table 5.

The native Fourier for gene *5* protein was calculated on planes xz using the program GENFOUR written by George Reeke at Rockefeller University. The map was contoured using a program written by Peter Campbell-Smith at Hershey and displayed on a Calcomp plotting device interfaced directly to a PDP 11/40 computer. The sections were then contoured onto acetate sheets and displayed in the conventional manner for visualization.

The electron density map of the gene *5* DNA unwinding protein is of good quality and from our experience compares favourably with those used to interpret a number of other macromolecular structures. One group of map sections containing the density of a strand of a β sheet running in the xz plane and a part of the β loop running more or less perpendicular to the sections is shown in Fig. 5 as an example of the map quality. The density map was interpreted in terms of the polypeptide backbone by examination of 9 in × 9 in mini maps. These were then photographed, projected and drawn onto 1 m × 1 m Mylar sheets and these used in a Richards Optical comparator (Richards, 1968) to construct a trial model of the structure with Kendrew model parts.

TABLE 1

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(trimethyl-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
$PtBr_2(NH_3)_2 + K_2Pt$-(trimethyldibenzyl 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

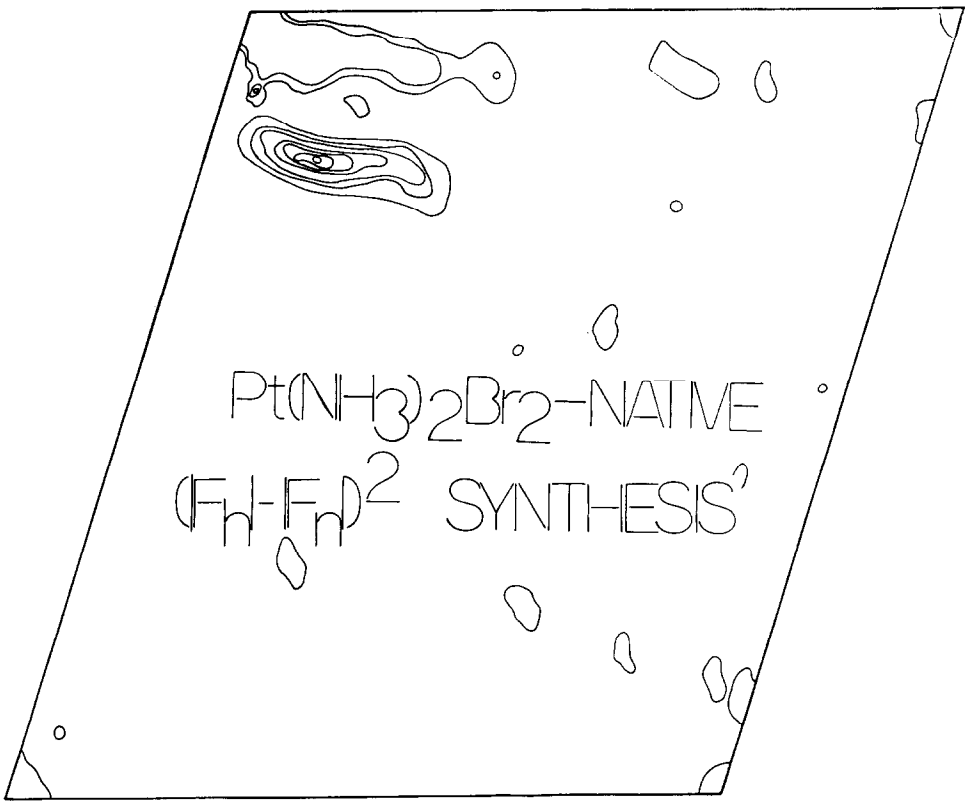

FIG. 2. Section $y = 0{\cdot}00$ of the difference Patterson synthesis of $Pt(NH_3)_2Br_2$-NATIVE GENE *5*, using coefficients $(|F_{Pt}| - |F_{NAT}|)^2$. The single large peak on the Harker section at $x = 0{\cdot}06$ $z = 0{\cdot}16$ is 3 times the weight of any other on the map.

3. Results and Discussion

The monomer, whose gross shape is seen in the wooden model of Figure 6, is essentially globular with a protruding appendage of density lying close to the molecular dyad. It is roughly 45 Å long, 25 Å wide and 30 Å high. The entire course of the polypeptide chain derived from our present 2·3 Å electron density map is shown in the stereo photographs of Figures 7 and 8. The molecular dimer which exists in solution must have the form shown in Figure 9. The known sequence of the protein has been fitted to the tracing of the polypeptide with no serious inconsistencies. There are three short regions of the chain which either contain portions of weak density or appear to show some measure of disorder. The bend of the β loop containing residues 22 to 27 is seen to be somewhat disperse. This bend, however, freely protrudes into the solvent-filled intersticies and might well be expected to show some positional flexibility even in the crystal. The other two sequences in question are 1 to 7 at the N terminus and 82 to 87 at the C terminus, both of which are partly in solvent regions. The N-terminal sequence appears to occupy either of two positions displaced slightly from one another. Those amino acids forming the C terminus simply show some degree of smearing out that we interpret as small statistical variations about a mean position. The tracing illustrated in Figures 7 and 8, therefore, is stated with a substantial degree of confidence.

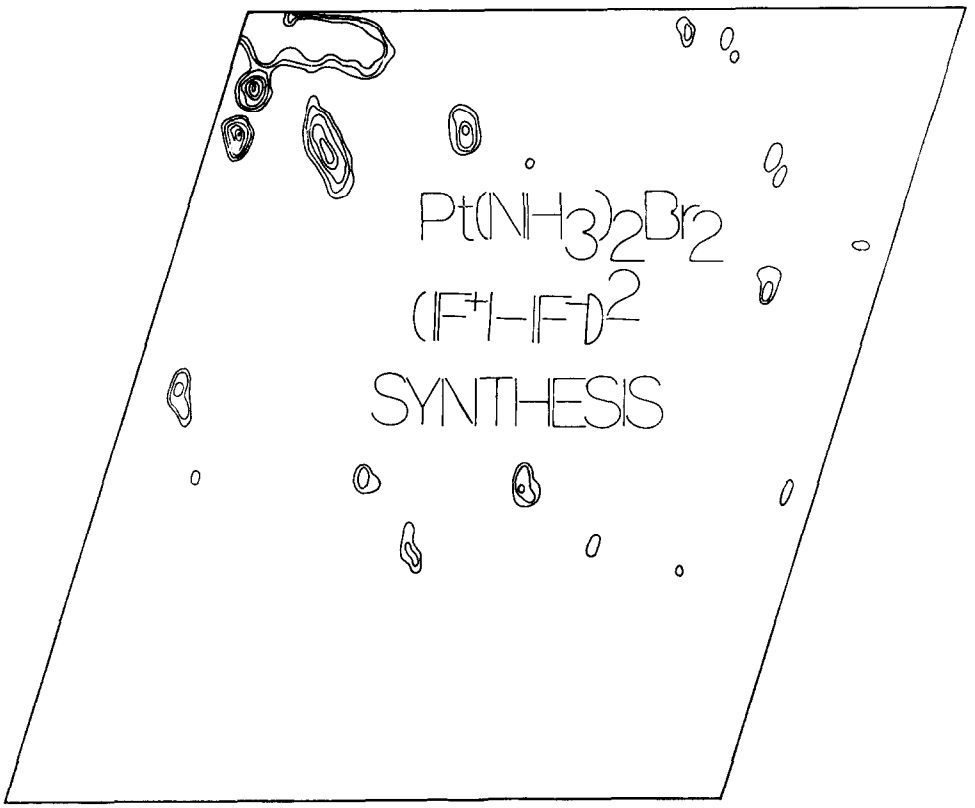

FIG. 3. Section $y = 0{\cdot}00$ of the anomalous Patterson synthesis of $Pt(NH_3)_2Br_2$ isomorphous derivative using coefficients $(|F^+_{pt}| - |F^-_{pt}|)^2$. The large peak at $x = 0{\cdot}06$, $z = 0{\cdot}16$ correlates exactly with that found in the isomorphous difference Patterson.

The protein is composed entirely of antiparallel β structure and short lengths of extended chain with no α-helix whatsoever. This is consistant with predictions based on physical-chemical data (Day, 1973) and sequence-structure rules (Anderson *et al.*, 1975; Coleman *et al.*, 1976). There are three basic elements of secondary structure that form the framework of the molecule, a three-stranded antiparallel β sheet (I) arising from residues 12 to 49, a two-stranded antiparallel β ribbon (II) formed by residues 50 to 70, and a second two-stranded antiparallel β ribbon (III) derived from residues 71 to 82. The arrangement of these elements as they occur in tracing from NH_2 to COOH terminus, and presumably the order in which they appear as the protein folds after synthesis, is shown in the series of drawings in Figures 10 and 11. The three-stranded sheet (I) emphasized in Figure 12 is not at all a flat surface but, as characteristic of small β sheets with short strands, is radically twisted. This severe distortion of the sheet produces a concavity or corridor along the underside of the molecule and is seen in the gross density as an overhanging ledge or archway. The β ribbons (II, III) emphasized in Figure 13 are also extremely twisted in character so that a precise assignment of the hydrogen bonding scheme may be somewhat tenuous and must await refinement of the atomic positions.

The β ribbon (II) lying close to the molecular dyad and at an angle of about 120° to the β sheet has two effects on the distribution of density. First, it enhances the

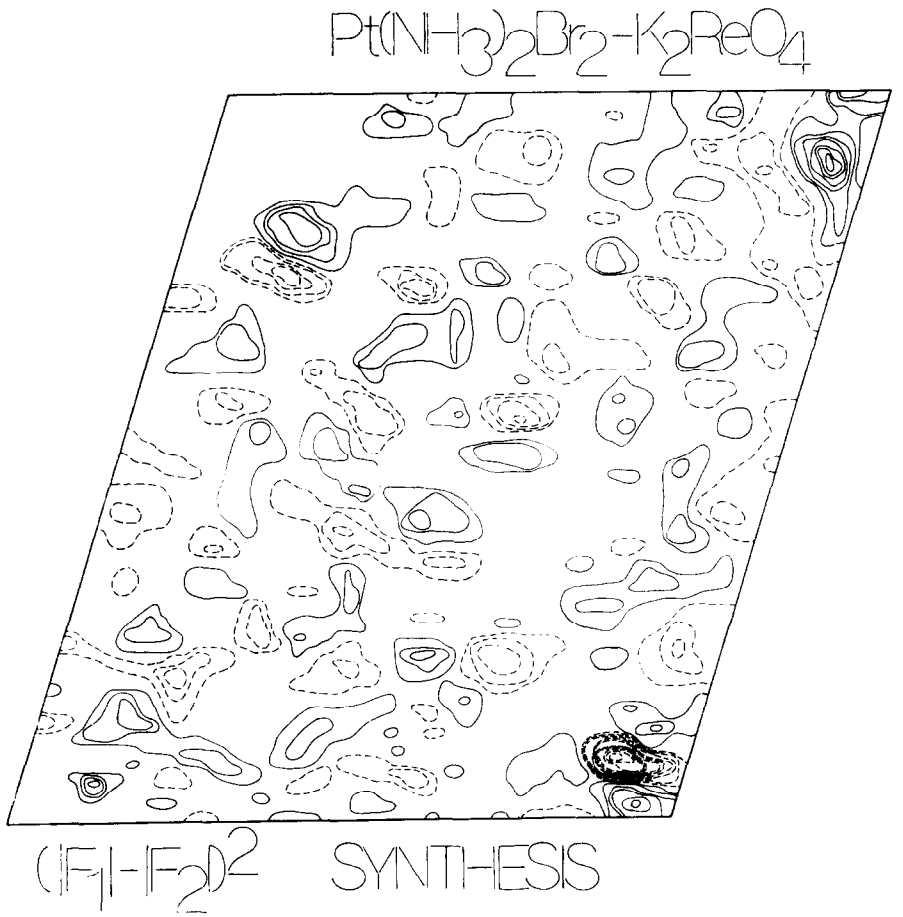

FIG. 4. Section $y = 0{\cdot}00$ of the difference Patterson synthesis of $Pt(NH_3)_2Br_2$-K_2ReO_4 derivatives using as coefficients $(|F_{Pt}| - |F_{Re}|)^2$. The large negative peak near the lower right corner corresponds exactly to the vector between the platinum site and the rhenium position on the dyad axis. This peak, approximately 3 times more negative than any other on the map serves to confirm both the platinum and rhenium co-ordinates.

tunnel-like nature of the concavity running beneath the sheet by creating a cradle at the point where the β ribbon most closely passes that of the β sheet. In addition, it results in a deep indentation in the monomer at the point where the interior edge of the β ribbon passes closest to the edge of the β sheet. The long groove beneath the three-stranded β sheet (I), by its shape and extent (~30 Å) would suggest it to be the DNA binding region. There is no other passage through the density that appears to be consistent with a long polynucleotide binding site. This contention is supported by the particular amino acid side-chains that form its surface. These include tyrosine 26, phenylalanine 13, cysteine 32 and tyrosines 41 and 56, all of which have been implicated in the binding of DNA by nuclear magnetic resonance and chemical modification studies (Coleman *et al.*, 1976; Anderson *et al.*, 1975; Day, 1973; Pretorius *et al.*, 1975). In addition, the sheet is comprised of residues from the NH_2-terminal half of the polypeptide, again the portion of the sequence implicated in binding by these studies (Coleman *et al.*, 1976).

Tyrosine 41, tyrosine 34 and phenylalanine 13 appear to stack upon one another at

TABLE 2

Co-ordinates 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	2·3 Å	0·0614
Iodine	0·2770	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 = 0·73.
Total number of reflections to 2·3 Å = 3793.

one end of the binding groove and look to be in position to interact with two consecutive bases of the bound DNA. Tyrosine 56 projects into the cradle region and is in position to stack upon bases of the polynucleotides as well. Tyrosine 26, on the other hand, is at the other end of the groove and is not paired with any other aromatic residues. There do not appear to be any additional aromatic side-chains extending into the binding region.

The single cysteine residue 33 is a part of the second strand of the β sheet (I) and therefore is at the edge of the molecule and part of the DNA binding groove. The side-chain, however, is turned up inside the density and sequestered in this conformation of the molecule. The adjacent tyrosine 34 is protected rather well from the solvent by other residues consistent with its incapacity to be chemically modified (Anderson *et al.*, 1975). It does not appear to be in close contact with the sulphur of cysteine 33 but does seem to be interleaved between tyrosine 41 and phenylalanine 13.

Our tentative assignment of the initial two thirds of the polypeptide chain was confirmed by the location of the major iodine binding site. This falls less than two angstroms from our choice for tyrosine 26, reported by Coleman *et al.* (1976) to be one of the three exposed tyrosine residues. Indeed it does occur on the underside of the three-stranded sheet (III) near the turn involving strands 1 and 2 and it is fully exposed.

TABLE 3

Refinement statistics for isomorphous derivatives of gene 5 protein

		$Pt(Br)_2(NH_3)_2$										
Zone number	Overall	1	2	3	4	5	6	7	8	9	10	11
Zone spacing *A*		7·70	5·36	4·73	4·23	3·68	3·32	3·10	2·88	2·69	2·51	2·38
No. per zone		244	145	130	239	412	260	284	466	330	544	258
RMS closure		123	77	76	76	58	53	50	54	41	50	50
RMS small F		177	156	140	140	125	103	92	90	82	68	58
RMS differences		179	133	117	114	93	79	73	71	65	70	70
RMS est. error		20	22	23	25	25	24	23	21	20	23	24
R values for general reflections												
Kraut *R* value	0·110	0·171	0·112	0·097	0·094	0·087	0·089	0·098	0·117	0·103	0·129	0·126
R modulus	0·481	0·562	0·429	0·472	0·442	0·393	0·436	0·463	0·478	0·418	0·586	0·715
R weighted	0·326	0·487	0·242	0·282	0·267	0·205	0·242	0·295	0·362	0·250	0·535	0·733
R values for centric reflections												
Kraut *R* value	0·176	0·230	0·221	0·120	0·129	0·142	0·132	0·168	0·245	0·186	0·185	0·128
Centric *R* value	0·671	0·643	0·589	0·595	0·457	0·609	0·810	0·793	0·913	0·766	0·876	0·762
R modulus	0·757	0·837	0·627	0·649	0·527	0·653	0·611	0·794	1·075	0·813	1·058	0·930

		$Pt(Br_2)(NH_3)_2$ + Pt-trimethyl, benzylamine										
Zone number	Overall	1	2	3	4	5	6	7	8	9	10	11
Zone spacing *A*		7·68	5·35	4·74	4·23	3·68	3·32	3·10	2·88	2·69	2·51	2·38
No. per zone		235	140	130	239	405	255	280	460	340	530	256
RMS closure		143	97	58	78	70	67	53	45	44	44	52
RMS small F		147	135	125	117	104	86	73	70	65	55	46
RMS differences		190	130	112	112	91	79	68	60	55	58	64
RMS est. error		23	24	27	27	27	27	25	26	27	29	30
R values for general reflections												
Kraut *R* value	0·120	0·207	0·139	0·091	0·105	0·109	0·118	0·112	0·100	0·110	0·109	0·131
R modulus	0·637	0·833	0·638	0·490	0·583	0·582	0·672	0·653	0·523	0·561	0·635	0·948
R weighted	0·525	0·944	0·529	0·272	0·410	0·440	0·552	0·517	0·396	0·458	0·637	1·294
R values for centric reflections												
Kraut *R* value	0·186	0·295	0·267	0·122	0·123	0·145	0·205	0·163	0·163	0·195	0·144	0·171
Centric *R* value	0·870	0·852	0·944	0·762	0·585	0·743	1·260	0·855	1·155	1·241	0·818	1·185
R modulus	0·954	1·191	1·044	0·739	0·520	0·784	1·225	0·970	1·168	1·029	1·025	1·343

K_2ReO_4

Zone number	Overall	1	2	3	4	5	6	7	8	9	10	11
Zone spacing A		7·80	5·35	4·74	4·23	3·68	3·32	3·10	2·88	2·69	2·52	2·38
No. per zone		245	146	131	242	418	262	288	455	344	529	271
RMS closure		59	35	53	51	45	59	47	52	61	45	41
RMS small F		158	135	125	115	99	85	78	67	59	50	42
RMS differences		116	100	107	102	86	75	60	58	63	55	51
RMS est. error		25	27	31	31	28	26	23	21	26	30	31
A. D. residuals		0·71	1·37	1·18	1·27	1·69	1·70	1·87	2·31	3·98	6·39	8·41
R values for general reflections												
Kraut R value	0·087	0·072	0·049	0·058	0·063	0·069	0·096	0·089	0·113	0·142	0·105	0·095
R modulus	0·415	0·229	0·196	0·286	0·330	0·355	0·497	0·423	0·527	0·724	0·580	0·632
R weighted	0·302	0·115	0·067	0·155	0·184	0·203	0·429	0·327	0·535	0·991	0·739	0·926
R values for centric reflections												
Kraut R value	0·135	0·142	0·101	0·096	0·126	0·095	0·134	0·122	0·179	0·214	0·198	0·151
Centric R value	0·604	0·436	0·357	0·537	0·524	0·538	0·682	0·779	0·903	1·087	0·894	1·208
R modulus	0·592	0·401	0·327	0·520	0·634	0·518	0·677	0·591	0·944	1·285	0·990	0·978

$Pt(Br)_2(NH_3)_2 + K_2ReO_4$

Zone number	Overall	1	2	3	4	5	6	7	8	9	10	11
Zone spacing A		7·73	5·35	4·74	4·23	3·68	3·32	3·10	2·88	2·69	2·51	2·39
No. per zone		226	139	128	230	402	245	249	377	260	391	152
RMS closure		114	73	87	80	68	53	58	49	44	40	41
RMS small F		137	117	111	104	90	79	75	63	56	51	43
RMS differences		158	108	120	107	91	75	77	64	57	52	54
RMS est. error		26	28	33	33	31	30	28	28	28	31	31
A. D. residuals		1·83	2·62	2·66	2·52	2·93	4·13	4·18	5·09	5·46	6·42	7·79
R values for general reflections												
Kraut R value	0·111	0·163	0·103	0·106	0·102	0·104	0·091	0·110	0·116	0·108	0·108	0·107
R modulus	0·622	0·652	0·500	0·628	0·622	0·625	0·564	0·623	0·641	0·622	0·641	0·800
R weighted	0·548	0·659	0·368	0·532	0·544	0·544	0·440	0·580	0·548	0·559	0·590	0·897
R values for centric reflections												
Kraut R value	0·167	0·251	0·184	0·103	0·180	0·136	0·102	0·183	0·154	0·111	0·180	0·104
Centric R value	0·716	0·714	0·614	0·474	0·683	0·793	0·710	0·919	0·927	0·696	0·953	0·780
R modulus	0·962	1·090	0·877	0·619	0·917	0·848	0·720	1·352	1·196	0·984	1·699	0·680

TABLE 3—*continued*

Zone number	Overall	Pt-trimethyl, benzylamine 1	2	3	4	5	6	7	8	9	10	11
Zone spacing A		7·56	5·35	4·73	4·23	3·68	3·32	3·10	2·88	0·0	0·0	0·0
No. per zone		219	138	127	230	391	238	259	430	0	0	0
RMS closure		129	90	83	83	70	73	72	62	0	0	0
RMS small F		105	100	99	92	80	66	70	64	0	0	0
RMS differences		147	102	102	99	91	84	87	72	0	0	0
RMS est. error		20	23	26	27	28	29	31	29	0	0	0
A. D. residuals		3·52	3·23	4·23	4·68	4·92	6·27	6·47	8·00	0·0	0·0	0·0
R values for general reflections												
Kraut *R* value	0·126	0·187	0·139	0·097	0·103	0·106	0·120	0·130	0·139	0·0	0·0	0·0
R modulus	0·867	1·047	0·845	0·675	0·775	0·795	1·000	0·894	0·874	0·0	0·0	0·0
R weighted	0·895	1·493	0·825	0·595	0·735	0·731	1·202	0·933	0·901	0·0	0·0	0·0
R values for centric reflections												
Kraut *R* value	0·158	0·230	0·203	0·091	0·140	0·116	0·174	0·158	0·156	0·0	0·0	0·0
Centric *R* value	0·846	0·956	0·891	0·786	0·768	0·719	0·849	0·824	0·918	0·0	0·0	0·0
R modulus	1·137	1·364	1·069	0·798	0·919	1·090	1·899	1·060	1·015	0·0	0·0	0·0

Zone number	Overall	Iodine 1	2	3	4	5	6	7	8	9	10	11
Zone spacing Å		7·79	5·35	4·73	4·24	3·68	3·32	3·14	0·0	0·0	0·0	0·0
No. per zone		235	138	127	223	391	218	124	0	0	0	0
RMS closure		79	85	103	110	113	113	123	0	0	0	0
RMS small F		106	101	100	99	93	89	93	0	0	0	0
RMS differences		117	114	118	136	131	136	127	0	0	0	0
RMS est. error		21	25	27	29	32	34	34	0	0	0	0
A. D. residuals		0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0	0·0
R values for general reflections												
Kraut *R* value	0·151	0·119	0·118	0·126	0·137	0·165	0·182	0·219	0·0	0·0	0·00	0·0
R modulus	0·928	0·631	0·706	0·874	0·964	1·056	1·093	1·142	0·0	0·0	0·0	0·0
R weighted	1·060	0·521	0·724	0·981	1·194	1·417	1·545	1·763	0·0	0·0	0·0	0·0
R values for centric reflections												
Kraut *R* value	0·204	0·184	0·154	0·156	0·211	0·233	0·204	0·413	0·0	0·0	0·0	0·0
Centric *R* value	0·776	0·634	0·658	1·013	0·745	0·883	0·860	0·894	0·0	0·0	0·0	0·0
R modulus	1·223	0·966	0·723	1·387	1·227	1·585	1·446	1·789	0·0	0·0	0·00	0·0

TABLE 4

Distribution of terms as a function of figure of merit

Figure of merit	0·05	0·15	0·25	0·35	0·45	0·55	0·65	0·75	0·85	0·95
No. of F's	45	72	135	154	226	278	382	486	726	977

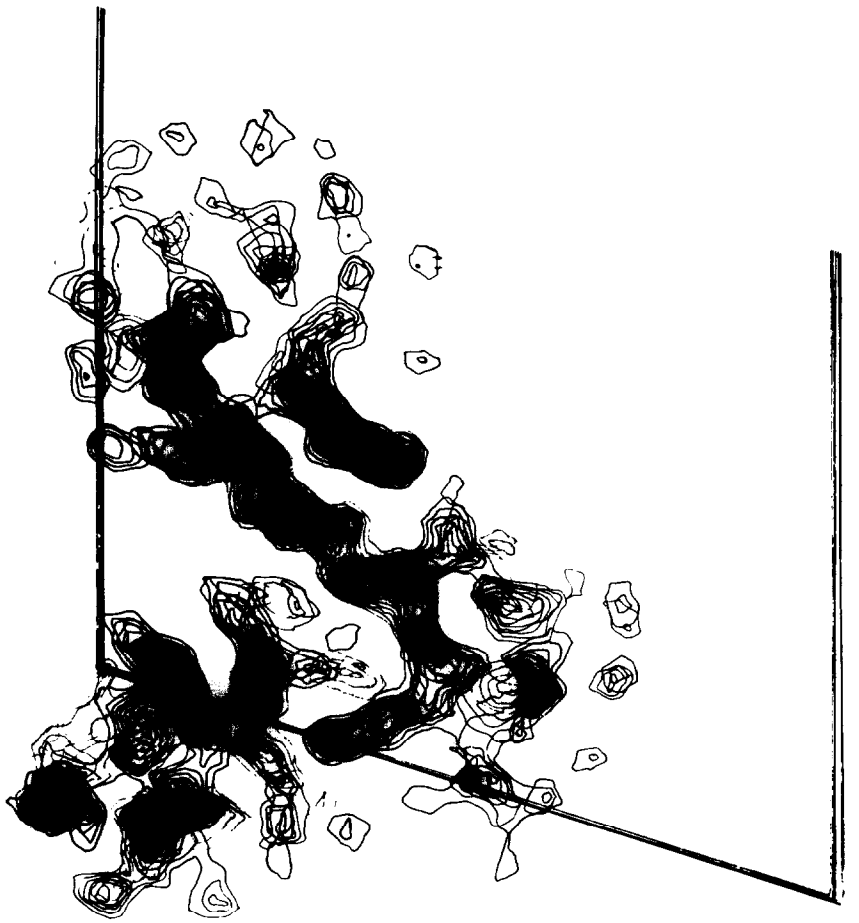

FIG. 5. Sections $y = 0{\cdot}07$ through $y = 0{\cdot}25$ of the electron density map of the gene 5 DNA unwinding protein projected onto a single plane. Strand 3 of the β sheet is seen lying almost in the plane of these sections as well as β loop (II) in the lower left which is running more or less diagonally through the sections.

The minor iodine site is very close to the major rhenium site and does not fall near a tyrosine residue. Because ReO_4^- also binds there, this site may represent a non-specific anion binding site in the molecule. Alternatively, this site is very near histidine 64, and this amino acid can also be iodinated.

The second major concavity in the electron density mass at the angle between the β ribbon (II) and the three-stranded β sheet (I) and seen best in Figure 7 is not an extended groove, and is furthermore not a candidate in the dimer for DNA binding.

TABLE 5

Figure of merit and residual distribution as a function of resolution

Zone number	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	0·475	0·457	0·328	0·365	0·423	0·428	0·542	0·516	0·499	0·474	0·614	0·933
Fig. of 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	265	491	357	578	298

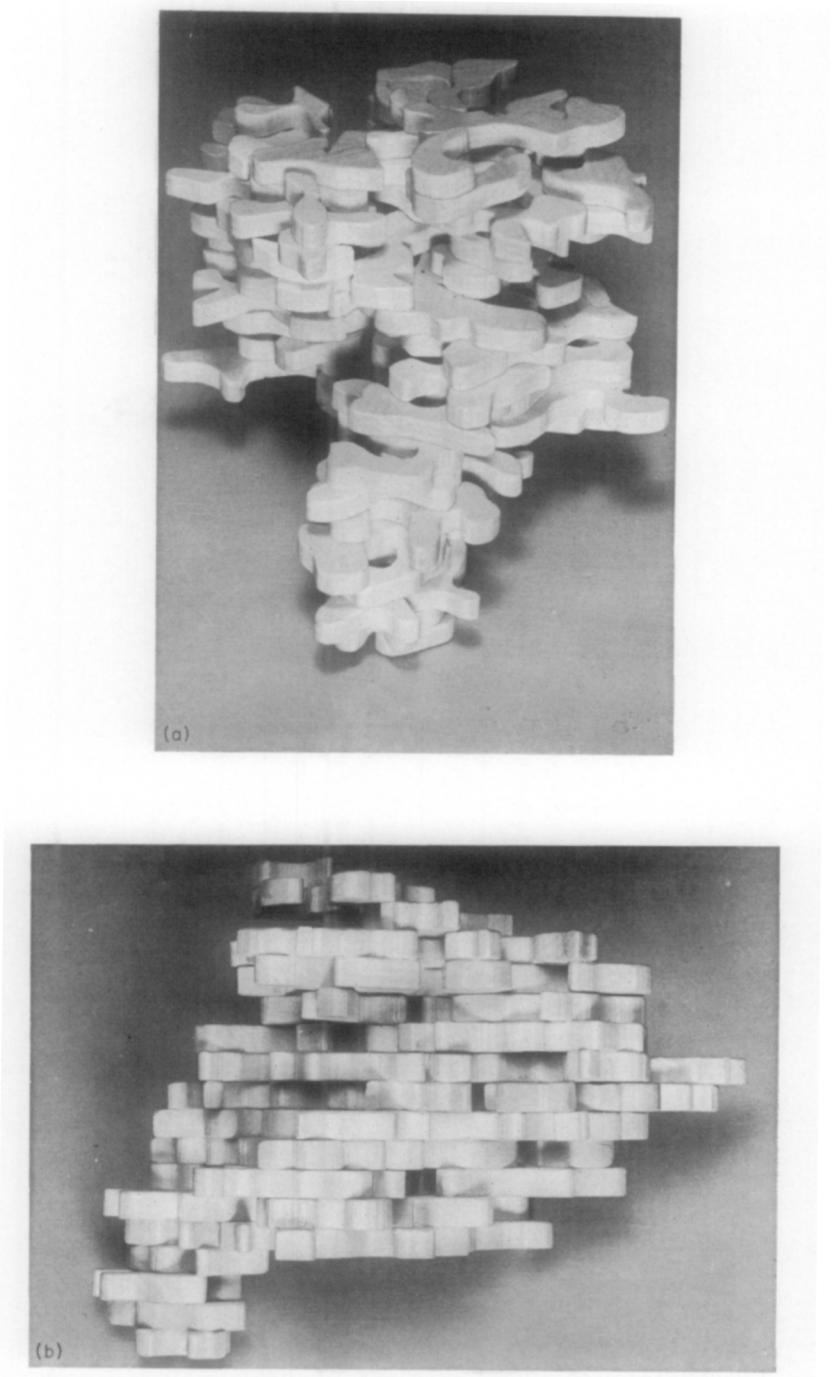

FIG. 6. A representation of the gene *5* electron density based on the 2·3 Å Fourier made by cutting appropriate envelopes of density from each section of map and assembling them in the y directions. In (a) the model is viewed approximately down the 011 direction and can be seen as an essentially globular mass with a protrusion of density extending below and away. In (b) the model is seen approximately down x. The presumed DNA binding region lies beneath the main body of density and inside the cradle formed by the density protruding to the upper right.

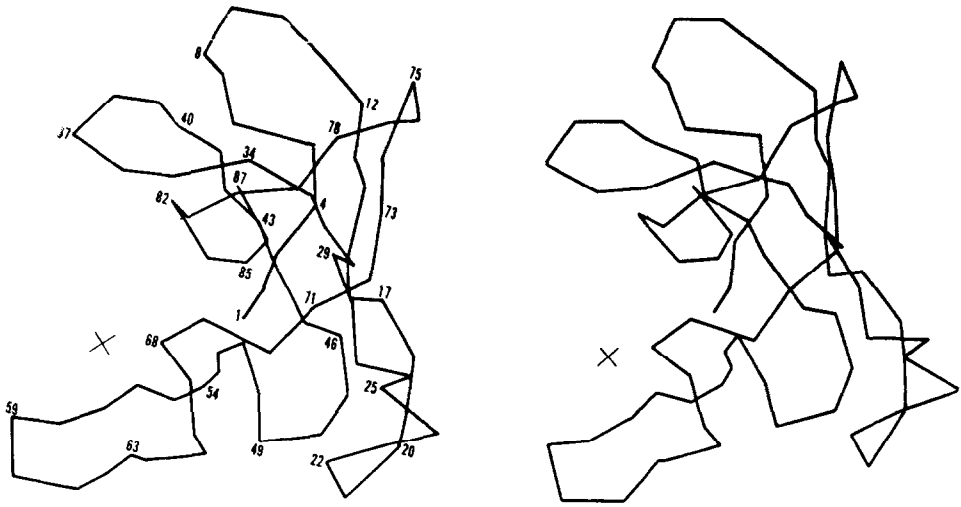

FIG. 7. Stereo photograph of the polypeptide backbone based on the α-carbon co-ordinates measured from the Kendrew model. The view is along the crystallographic *a* axis.

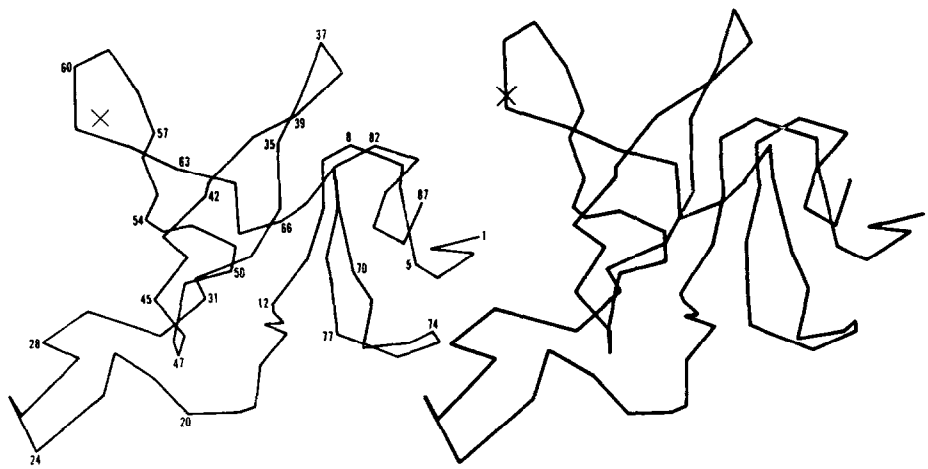

FIG. 8. Stereo photograph as in Fig. 7 but the view is along *b*.

Because of the molecular dyad relating the two monomers in the dimer, the cavity is entirely filled by the symmetry related density of the β ribbon (II′). This symmetrical association between two dyad related loops II and II′ is undoubtedly the interaction that maintains the dimer as the predominant species in solution. The actual contact area would seem to involve, at least to some extent, a large number of hydrophobic residues, although other kinds of side-chains are present as well. The amino acid sequence 54 to 59, Pro-Ala-Tyr-Ala-Pro-Gly, forms part of one strand of loop II closest to the dyad. Since this sequence would not be expected to readily form β structure it is unclear how much β character to ascribe to loop II. Loop II also contains the single histidine residue 64 which seems to have the imidazole group directed back into the central density of the loop. It too is unreactive in the crystal toward heavy metal compounds.

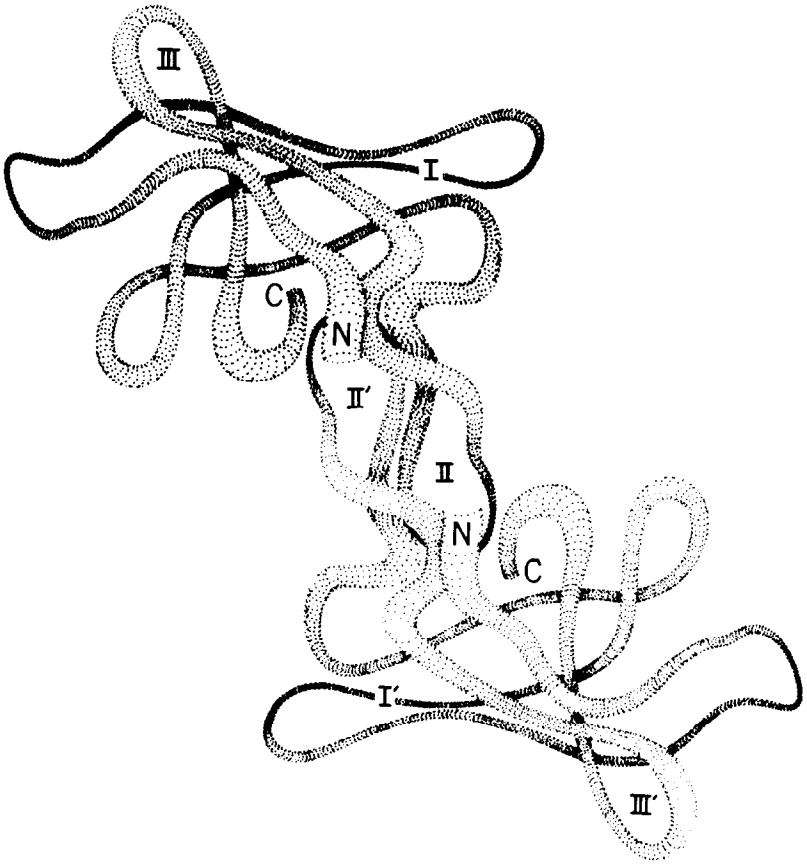

FIG. 9. A drawing of the 2 gene *5* polypeptide chains that comprise the dimer showing their association about the molecular dyad through interaction of the symmetry related β loops II and II′.

Coleman *et al.* (1976) pointed out that the gene *5* monomer contains a high proportion of hydrophobic residues and, because of its small size, an appreciable number of these must be exposed to solvent. The dimer interaction, however, will cause many of these hydrophobic residues to be shielded from solvent that would otherwise be exposed. This may be part of the driving force that maintains the dimer in solution. Since additional residues will likely be shielded from solvent by lateral interactions when binding occurs along the DNA strands, this same mechanism may also be largely responsible for the highly co-operative nature of the protein binding.

The second β ribbon (III) lies almost directly above and diagonally across the three-stranded sheet (I); it is, however, separated for the most part from the DNA binding region by the sheet. The bend is formed from residues 75 to 78, Ser-Leu-Met-Ile. The non-terminal methionine, although apparently accessible to solvent, could not be reacted with heavy metal reagents in the crystal. This seems to result from the close juxtaposition of the bend with electron density from a second molecule in the unit cell with concomitant obstruction to small compounds. We believe that the β loop III and the residues of the carboxy-terminal region could be primarily involved

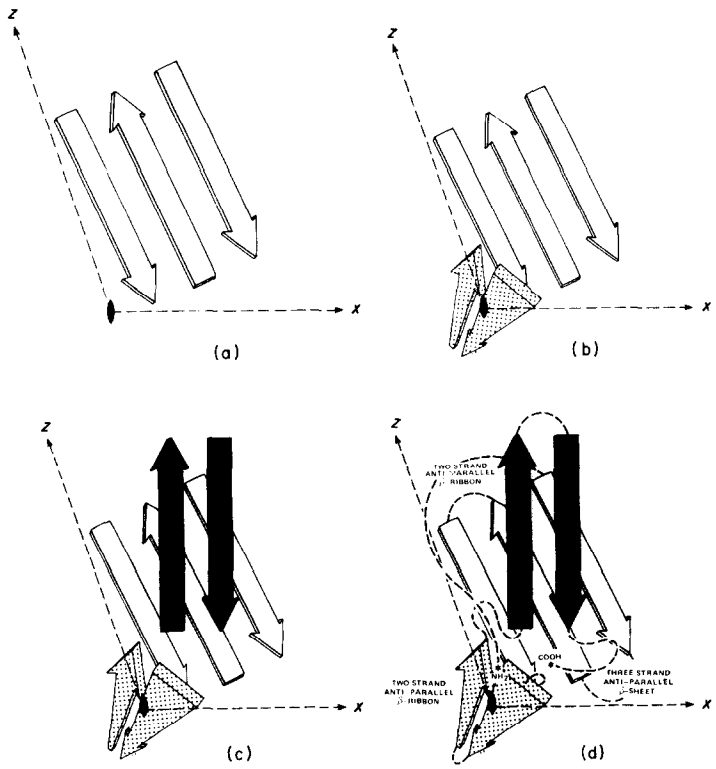

FIG. 10. A series of 4 schematic drawings showing the development of the secondary structural components of the gene 5 molecule from the NH_3-terminal methionine to the COOH-terminal lysine. In (a) the 3 strands of the antiparallel β sheet (I) which form the major part of the DNA binding region arise from the amino-terminal half of the sequence. In (b) the 2 strands of the β ribbon (II) appear. This element appears to be principally responsible for maintaining the molecule as a dimer in solution by its interaction with a symmetry related β ribbon. In (c) the second β ribbon (III) is established diagonally across the β sheet (I). We suspect this component may be the primary participant in the lateral interactions from which the co-operativity of the DNA binding arises. In (d) the segments of extended polypeptide chain connecting the secondary structural elements are indicated as well as the extended amino and carboxy-terminal portions of the backbone.

in the protein–protein interactions responsible for the co-operative, contiguous binding along the DNA strands.

The final six residues at the COOH terminus are in extended conformation but coiled back into the central mass of the molecule. The electron density corresponding to these amino acids is the most difficult in the map to interpret and suggests some disorder in this part of the polypeptide chain. The initial ten residues of the polypeptide chain are also in extended conformation and wind from the amino-terminal methionine over the top of the molecule and enter strand 1 of the β sheet (I). The sulphur atom of the methionine 1 lies only about 5 Å from the molecular dyad and is the residue responsible for binding the platinum heavy-atom derivatives. The electron density map shows that the N terminus occupies either of two positions which are separated by about 5 Å. Thus we see in the map two large density peaks for the sulphur but in each case they are joined back to the polypeptide chain at lysine 3 by two separate branches of density assigned to the two alternate positions for Ile2. This

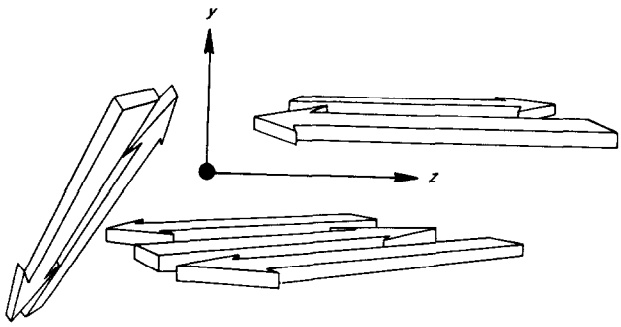

FIG. 11. A schematic drawing of the gene *5* secondary structural elements arranged in space when viewed along the x direction.

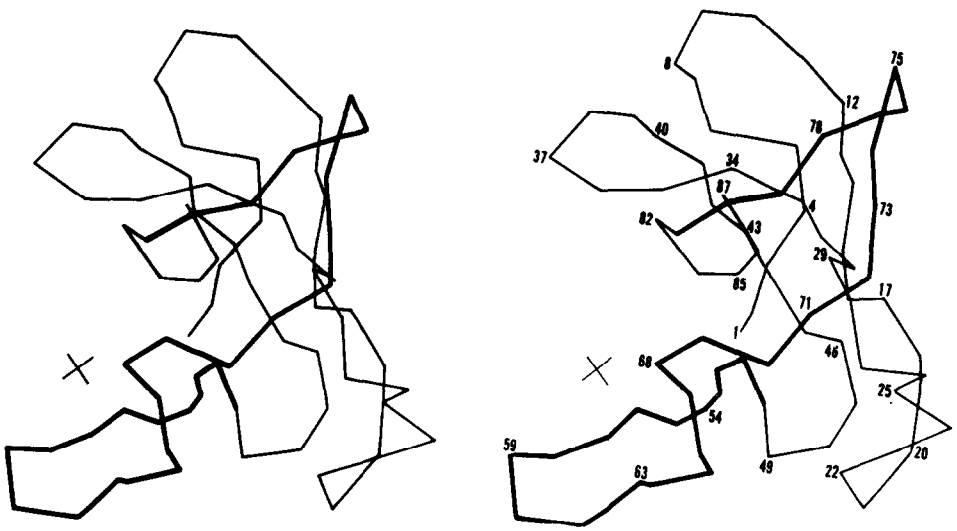

FIG. 12. Stereophotograph of the polypeptide backbone as in Fig. 8 with the 3-stranded β sheet (I) emphasized.

disorder caused by two alternate positions in the crystal for the first three residues also explains the secondary sites for the platinum derivatives.

The gene *5* protein would seem, in structure-function terms, to be a very economical molecule, and we believe we can tentatively assign one of its three major functional responsibilities to each of its three prominent secondary structural elements. The three-stranded antiparallel β sheet (I) is primarily responsible for binding to and interacting with the DNA. The two-stranded antiparallel ribbon (II) maintains the dimer in solution by tightly interlocking with a symmetry related loop (II'). The second β ribbon (III) we believe to be involved in the co-operative protein interactions with neighbouring molecules. Thus, it would appear that in the gene *5* protein, we find the minimum amount of structure expended to fulfil the required functional demands.

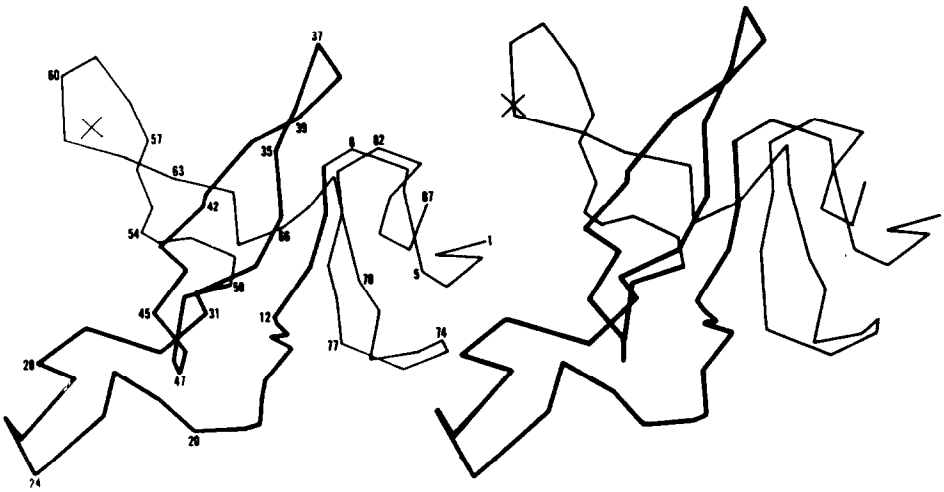

FIG. 13. Stereophotograph of the polypeptide backbone as in Fig. 7 with the 2 antiparallel loops II and III emphasized.

The molecular packing is such that there are some extensive solvent regions separating dimers from each other and channels through the crystal as well. The molecules, however, are stacked very closely along the y direction by both translation and as a result of the 2_1 screw axis at $x = 1/4$, $z = 1/2$. This has the net effect of filling or blocking the presumptive DNA binding groove with density from neighbouring molecules, thereby explaining our inability to bind short oligonucleotides, from one to four in length, to the protein as it exists in these crystals. Thus it does not appear that we will be able to determine the structure of gene 5–oligonucleotide complexes by simple diffusion and difference Fourier synthesis.

This research was supported by grants from the National Institutes of Health, National Science Foundation, and the American Cancer Society. One of the authors (A. H. J. W.) is suported by a grant from the Massachusetts Institute of Technology Cancer Center. Another of the authors (F. A. J.) is a National Institutes of Health postdoctoral fellow.

REFERENCES

Adams, M. J., Hass, D. J., Jeffrey, B. A., McPherson, A., Mermall, H. L., Rossmann, M. G., Schevitz, R. W. & Wonacott, A. J. (1969). *J. Mol. Biol.* **41**, 159.

Alberts, B., Frey, L. & Delius, H. (1972). *J. Mol. Biol.* **68**, 139–152.

Anderson, R. A., Nakashima, Y. & Coleman, J. E. (1975). *Biochemistry*, **14**, 907.

Arndt, U. W. & Willis, B. T. M. (1966). *Single Crystal Diffractometry*, chapter 10, University Printing House, Cambridge, U.K.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Crystallogr.* **12**, 794.

Cavalieri, S., Goldthwait, D. A. & Neet, K. (1976). *J. Mol. Biol.* **102**, 713.

Coleman, J. E., Anderson, R. A., Ratcliffe, R. G. & Armitage, I. M. (1976). *Biochemistry*, **15**, 5419.

Day, L. A. (1973). *Biochemistry*, **12**, 5329.

Dickerson, R. E., Kendrew, J. C. & Strandborg, B. E. (1961). *Acta Crystallogr.* **14**, 1188.

Dunker, A. K. & Anderson, E. A. (1975). *Biochim. Biophys. Acta*, **402**, 31–34.

Laemmli, V. K. (1970). *Nature (London)*, **227**, 680.

McPherson, A. (1976*a*). In *Methods of Biochemical Analysis* (Glick, D., ed.), vol. 23, pp. 249, Wiley and Sons, New York.

McPherson, A. (1976*b*). *J. Biol. Chem.* **251**, 6300.

McPherson, A., Molineux, I. & Rich, A. (1976). *J. Mol. Biol.* **106**, 1077.
Nakashima, Y., Dunker, A. K., Marion, D. A. & Konigsberg, W. (1974). *FEBS Letters*, **40**, 290.
North, A. C. T. (1965). *Acta Crystallogr.* **18**, 212.
Oey, J. L. & Knippers, R. (1972). *J. Mol. Biol.* **68**, 125–128.
Pratt, D., Laws, P. & Griffity, J. (1974). *J. Mol. Biol.* **82**, 425–439.
Pretorius, H. T., Klein, M. & Day, L. A. (1975). *J. Biol. Chem.* **250**, 9262–9269.
Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225.
Rossmann, M. G. (1960). *Acta Crystallogr.* **13**, 221.
Salstrom, J. S. & Pratt, D. (1971). *J. Mol. Biol.* **61**, 489–501.
Wyckoff, H. W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L. N., Hardman, K. D., Allewell, N. M., Kelly, D. M. & Richards, F. M. (1967). *J. Mol. Biol.* **27**, 563.