THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 255, No. 7, Issue of April 10, pp. 3174-3177, 1980
Printed in U.S.A.

X-ray Diffraction Studies on Crystalline Complexes of the Gene 5 DNA-unwinding Protein with Deoxyoligonucleotides*

(Received for publication, July 23, 1979, and in revised form, December 12, 1979)

Alexander McPherson,‡ Andrew H. J. Wang,§¶ Frances A. Jurnak,‡‖ Ian Molineux, Frank Kolpak,§‡‡ and Alexander Rich§**

*From the ‡ Department of Biological Chemistry, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033, § The Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and the ** Imperial Cancer Research Fund, Mill Hill Laboratories, London, England*

Complexes of the gene 5 protein with a variety of oligodeoxynucleotides, ranging in length from two to eight and having several different sequences, have been formed and crystallized for x-ray diffraction analysis. The crystallographic parameters of four different unit cells, all of which are based on hexagonal packing arrangements, indicate that the fundamental unit of the complex is composed of 12 gene 5 monomers.

The cumulative evidence acquired by molecular biologists over the past 20 years points to the interaction of protein molecules with DNA and RNA as a primary source of genetic and metabolic control in living organisms. A detailed description of the means by which proteins interact with the control points and transcribable genes of DNA will suggest the mechanisms by which the flow of information is mediated. Such a description does not seem possible unless the molecular components and distribution of chemical bonds can be directly visualized. The only available means for attaining such an image is through application of the x-ray diffraction technique to single crystals of DNA-binding proteins which form complexes with fragments of the nucleic acid. With these objectives in mind, we have attempted to obtain crystalline complexes of DNA-binding proteins with short fragments of the nucleic acid. We have succeeded in the case of the DNA-unwinding protein gene 5 product of fd bacteriophage, a protein which we have already crystallized and studied in the unliganded form (1).

Extracts of *E. coli* infected with a filamentous bacteriophage, such as fd or M13, when chromatographed on single-stranded DNA-cellulose yield on elution a protein of M_r = 9,800 which is made in at least 100,000 copies/cell (2). This protein, coded for by gene 5 of fd or M13 is absolutely required to direct the synthesis of progeny single-stranded DNA (3). By forming a stoichiometric complex with single-stranded DNA, it disturbs the equilibrium of SS⇌RF where SS is single strand DNA and RF is replicative form DNA. Consequently, the pool of replicative form DNA molecules continues to generate progeny single strand DNA which, when it forms a complex with the gene 5 protein, is unable to serve as a template for synthesis of the replicative form (4, 5).

The gene 5 product of fd phage has 87 amino acid residues whose sequence has been determined (6). The molecular species in solution appears to be a dimer of 20,000 daltons (7). Each gene 5 protein monomer covers approximately four nucleotides along a DNA strand (2, 3, 8, 9). The protein is nonspecific in that it will bind to oligonucleotides containing any of the four common bases in any sequence, although the binding affinity may vary somewhat (10). The binding of gene 5 protein to single-stranded DNA is cooperative in that a free gene 5 molecule prefers to bind adjacent to a protein molecule already bound to the DNA rather than to an isolated site on the DNA (11). The additional cooperative binding affinity is presumably contributed by protein-protein interactions between adjacent dimers in the complex.

The gene 5 protein·DNA complexes produced *in vitro* and studied by electron microscopy are unique in that two protein-covered strands coalesce to yield a helical rod-like structure in which there are 12 gene 5 monomers/turn of the helix. The helix has a width of approximately 100 Å and a longitudinal repeat of about 70 Å (2). The gene 5 protein·DNA complexes resemble mature filamentous bacteriophage virions, although 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 (12, 13). The gene 5 protein is never found in the virion but is returned to the cell for reuse. The complexes isolated directly from infected cells (9) are apparently different from those formed *in vitro*. *In vivo* complexes are composed of a fiber 40 Å in width which is supercoiled to give an overall width of 160 Å and a longitudinal repeat of 160 Å.

Because of the high affinity the gene 5 protein demonstrates for oligonucleotides, the protein can serve as a workbench molecule for the study of protein-nucleic acid interactions. Extensive NMR studies of complexes of gene 5 protein with tetranucleotides and octanucleotides of both random and specific sequence have been carried out; a great deal of physical and chemical data has been accumulated and predictions have been made as to the manner by which the associations take place (8, 10, 11, 14, 15). Thus an x-ray crystallographic analysis of these gene 5 protein·DNA complexes will not only yield the dual molecular structure but also serve to test the validity of the physical and chemical approaches to structure analysis of these types of systems.

MATERIALS AND METHODS

The oligonucleotides used in the crystallization experiments were d-pGpC, d-pApT, $d(Ap)_4$ and $d(Ap)_8$ from Collaborative Research,

* This research was supported by grants from the National Institutes of Health, the National Science Foundation, the National Aeronautics and Space Administration, and the American Cancer Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Supported by a grant from the Massachusetts Institute of Technology Cancer Center.

‖ National Institutes of Health Postdoctoral Fellow.

‡‡ Fellow of the Anna Fuller Fund.

Waltham, Mass. The specific sequence oligomers d-pCpTpTpC and d-$(Tp)_4$ were gifts of Drs. Robert Ratliff and Lloyd Williams of the University of California at Los Alamos, and $d(Cp)_3$ and $d(Cp)_4$ were gifts of Dr. Gobind Khorana of the Massachusetts Institute of Technology. d(GGTAAT) and its complement were provided by Dr. Jack Van Boom of the University of Leiden. Gene 5 protein from fd bacteriophage was made by a slight modification of the procedure of Alberts *et al.* (2). Infected cells were lysed in a French Press in 0.02 M Tris (pH 7.6), 0.01 M $MgCl_2$, with no added DNA. The lysate then was made 2 M in NaCl and nucleic acids were removed by polyethylene glycol (PEG) precipitation. The supernatant was dialyzed to low salt by repeated changes of a dialysis buffer containing 0.02 M Tris (pH 7.6), 0.05 M NaCl, 0.005 M EDTA, 0.001 M 2-mercaptoethanol, and 10% (v/v) glycerol. The protein then was isolated by the procedure of Alberts *et al.* (2), using DNA-cellulose and DEAE-cellulose chromatography. Occasionally the gene 5 protein had to be purified further by phosphocellulose chromatography to yield a homogeneous sample.

Crystallization was by the vapor diffusion method using depression plates and the precipitating agent was PEG 6000[1] from Fisher (16, 17). The methods and conditions were virtually the same as those described for the native gene 5 protein crystallization (18). The samples contained 5 μl of 14 mg/ml of gene 5 protein, 5 μl of the oligonucleotide solution, both buffered at pH 7.5 by 0.01 M Tris-HCl, and 10 μl of 12% PEG 6000 in H_2O. The sandwich box reservoirs contained 25 ml of 12% PEG 6000. All operations were carried out at room temperature. The time required for crystal growth to occur varied from several days to several weeks with the mean at about 20 days. As described below, a variety of crystal forms were observed. The crystals in general were stable in their mother liquor for up to several months, although some were degraded apparently in response to temperature fluctuations.

In several cases where large masses of microcrystals or crystals otherwise unsuitable for x-ray diffraction analysis formed, the crystalline material was collected by microcentrifugation of several pooled samples. The microcystalline samples were thoroughly washed with 12% PEG 6000 and dissolved in water with the aid of small amounts of NH_4OH. The optical densities at 260 and 280 nm then were measured on a Zeiss spectrophotometer in a 1-ml sample cuvette.

For x-ray diffraction analysis, the crystals were mounted by conventional means in quartz x-ray capillaries along with a small aliquot of mother liquor and sealed with dental wax. X-ray diffraction photographs were recorded on Ilford Industrial G x-ray film, using a Buerger precession camera with a crystal-to-film distance of 90 mm and exposure times of about 24 h. Nickel-filtered CuK_α x-radiation was produced by an Elliott rotating anode generator operated at 40 kV and 40 mA with a focal spot size of 200 μm^2.

RESULTS

Attempts were made initially to form complexes of the gene 5 protein with mono-, di- and trinucleotides by diffusing the oligomers into native gene 5 protein crystals. Such complexes, it was hoped, would be sufficiently isomorphous with the native structure that they could be solved directly using difference Fourier methods. Unfortunately, we have been unsuccessful in these attempts, with all of the derivative crystals exhibiting diffraction intensities identical to those of native crystals. It is now clear from the electron density map of the native structure (1) why this is so. In the monoclinic gene 5 protein crystals, the molecules related by unit cell translation along *y* are packed closely so that the region we postulate to be the binding site for the DNA on one monomer is completely blocked by density from another monomer. Thus, the oligonucleotides probably enter the crystals but are unable to diffuse into the binding site.

A second approach was to co-crystallize the gene 5 protein in the presence of deoxyoligonucleotides in the hope that complexes would form and that those complexes then would crystallize. In all cases, we used oligomers of determinate length and specific sequence so as to achieve as high a degree of homogeneity as possible.

Several different crystal forms of the gene 5 protein which formed complexes with oligonucleotides have been observed under a variety of different crystallization conditions. Common among these are the hexagonal plates seen in Fig. 1, diamond plates as seen in Fig. 2, rhombohedra, hexagonal prisms, and trigonal prisms. The diamond plates are observed most frequently and seem to be the dominant habit. In several cases we have observed transformations between these different forms over time as well as polymorphism within a single sample of mother liquor.

We find there to be little correlation between the length or specific sequence of the oligonucleotide combined with the gene 5 protein and the particular crystal form observed. For example, the diamond plates have been grown from gene 5 protein which formed complexes with d(GpC), d(ApT), $d(Ap)_4$, $d(Ap)_8$, or d(pCpTpTpC). Similarly the hexagonal prisms have been grown from gene 5 protein which formed complexes with the hexamer as well as with its complement. We find that at oligonucleotide/protein ratios of less than 1.0 only native crystals are grown. At concentrations above this level only the complex crystals appear. We have in no case observed the coexistence of native crystals with complex crystals. Furthermore, we have never observed the complex crystals under any other sets of conditions in the absence of added oligonucleotide. We have not noted any upper limit on the ratio of oligonucleotide/protein so far as crystal growth is concerned but have not gone beyond about 8:1 with the higher oligomeric fragments.

In those experiments where microcrystalline material was washed and redissolved and the 260:280 nm absorbance ratio

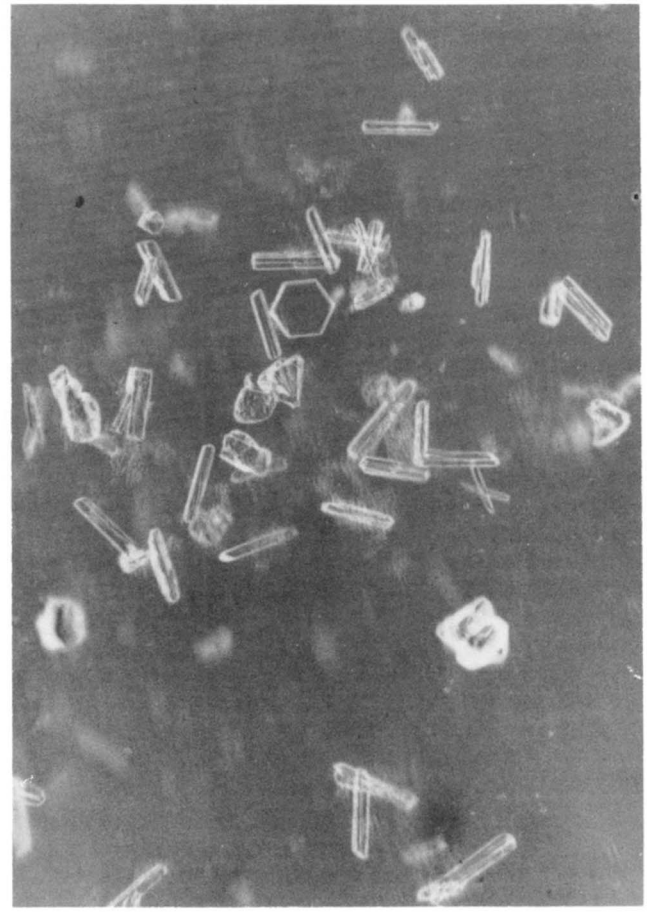

FIG. 1. **Low power light microscope photograph of the hexagonal plate crystal form of the complex between gene 5 protein and d-pCpTpTpC.** These have been observed to co-exist in a single sample with the diamond plates shown in Fig. 2. Their space group is $P6_3$ and they diffract to beyond 3.0 Å resolution.

[1] The abbreviation used is: PEG, polyethylene glycol.

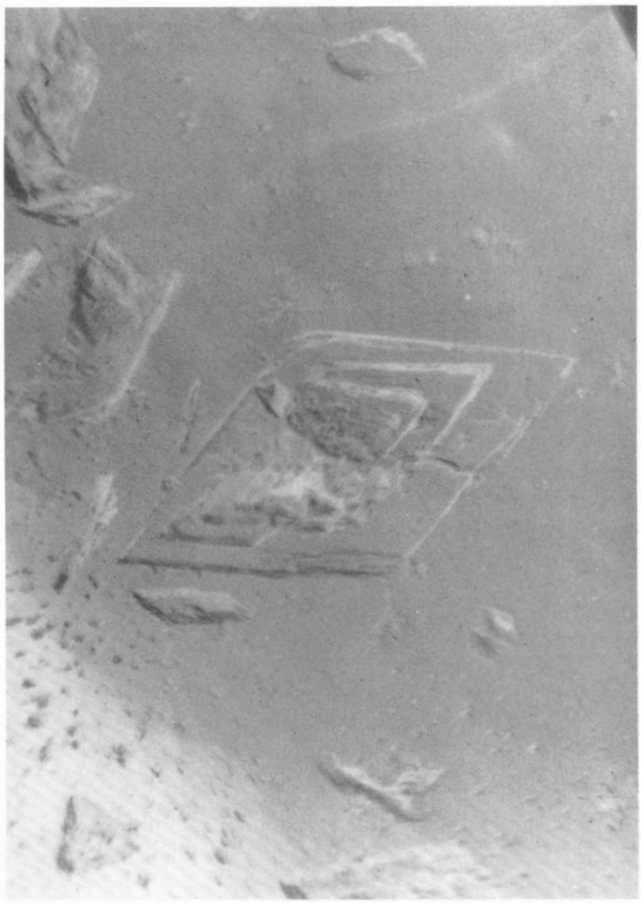

FIG. 2. **Low power light microscope photograph of the diamond plates which are the predominant form of gene 5 protein-oligonucleotide crystals.** These belong to space group $C222_1$ but can be related easily to the $P6_3$ hexagonal form seen in Fig. 1. The disorder apparent in these crystals has proven a serious problem in the x-ray analysis.

was measured, the results were consistent with the presence of between 0.75 and 1.0 oligonucleotide for each gene 5 monomer. The actual values of these ratios, or course, varied with the particular oligomer used to form the complex and these varied in length in these experiments from two to six. Although the accuracy of these measurements is not sufficient to establish an exact stoichiometry for the crystalline complexes, they do reassure us that nucleic acid is present in substantial amounts in the crystals.

Most of the crystals grown were multiples, frequently twinned, were of too small a size, or presented some problem with regard to x-ray diffraction analysis. Several of each form, however, grew to quite large size and permitted us to determine their space groups and cell dimensions by x-ray diffraction analysis. In some cases the limited extent of the diffraction patterns make our assignments somewhat uncertain, particularly with regard to systematic absences along axial rows where only a few orders could be reliably observed. Nevertheless, we believe we have correctly deduced the unit cell parameters for four crystal forms.

The hexagonal plates were among the best crystals in terms of physical strength. They are of space group $P6_3$ and have a = 107 Å and c = 206 Å. The unit cell volume is 2.04×10^6 and contains six crystallographic asymmetric units. Although we were unable to measure accurately the density of any of these crystals, we estimated the number of gene 5 protein molecules/asymmetric unit by assuming a volume/mass ratio (V_m) near the center of the range of those proteins described by Matthews (19). This is equivalent to assuming a density ρ for the crystals of about 1.15 g/cm^3 or a solvent content of 45%. This is approximately the same as that found for the native gene 5 protein crystals (18). We further assumed that the asymmetric unit is composed of dimers of gene 5 protein since that is the dominant species present in solution (7) and in the native crystal. With these assumptions, we calculate the number of gene 5 monomers/asymmetric unit to be 12 for a V_m of 2.83 Å^3/dalton. Thus, the crystallographic asymmetric unit has a molecular weight of 120,000.

The flat diamond plates gave diffraction intensities that fell on a rectangular nonprimitive net with real cell dimensions of a = 110 Å, b = 180 Å, and c = 117 Å. The space group, at the limit of our resolution (~10 Å), is $C222_1$, although the c axis could conceivably be a dyad if higher order odd 001 reflections are present. The volume of this orthorhombic unit cell is 2.32×10^6 Å^3 and it contains 8 asymmetric units. Using the same assumptions as above, the number of gene 5 protein molecules/asymmetric unit is again calculated to be 12, for a V_m of 2.42 Å^3/dalton.

The third crystal form studied, which appeared to be rhombohedral in habit, was that of space group R32. The rhombohedral cell dimensions are a = 140 Å with $\alpha \simeq 60°$. The dimensions for the equivalent triple hexagonal unit cell are a = 200 Å and c = 205 Å and its volume is 7.10×10^6 Å^3. The hexagonal cell must contain 18 asymmetric units/cell. Using the same assumptions once again, the number of gene 5 monomers/asymmetric unit is calculated to be 12.

The hexagonal prisms were the best diffracting crystal form we studied. They grew to substantial size and scattered to Bragg spacings of at least 3.0 Å in precession photographs, such as that seen in Fig. 3. The space group is trigonal $P3_1$ with a = 143 Å and c = 83 Å. The volume of 1.58×10^6 Å^3 must accommodate 3 asymmetric units and this requirement is best fulfilled if one assumes there to be 24 gene 5 monomers as the asymmetric unit. This implies a solvent content of about 40% and a V_m of about 2.2 Å^3/dalton. These trigonal crystals, however, show very high 32 pseudosymmetry in their

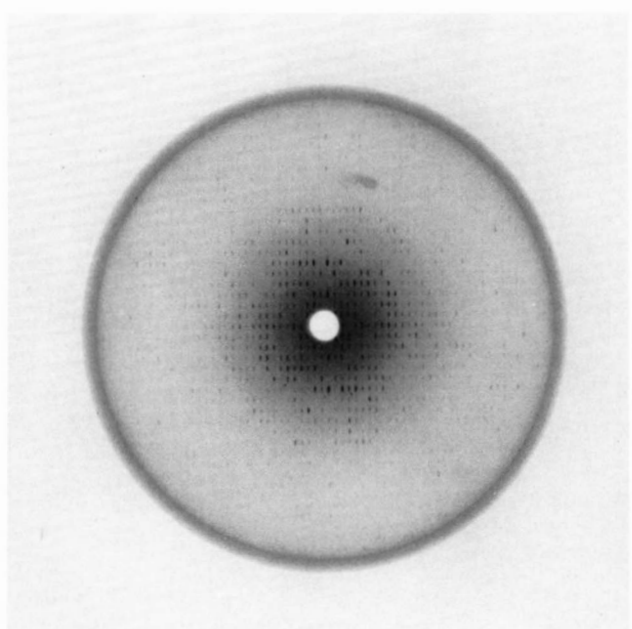

FIG. 3. **Precession x-ray diffraction photograph of the *h0l* zero level of reciprocal space of the $P3_121$ crystals composed of gene 5 protein which formed a complex with d(GGTAAT).** These crystals grow to large size and diffract strongly to at least 3.0 Å in precession photographs.

diffraction patterns and are very nearly of space group $P3_121$. Thus the pseudo-unit cell with the same V_m and solvent content would again have 12 gene 5 monomers as the asymmetric unit.

DISCUSSION

The diffraction patterns of at least two of the crystal forms grown extend to 3.0 Å resolution or better and seem to present an outstanding opportunity for the visualization of a protein·nucleic acid complex to nearly atomic resolution. The size of the asymmetric unit in all of the crystals suggests this to be a considerably more difficult crystallographic problem than that of the native gene 5 protein crystals. We do, however, believe that applying molecular replacement techniques with the native structure as the search model will elucidate the arrangement of the gene 5 protein dimers in the aggregate forming the asymmetric unit of the complex crystals. Indeed, we expect these techniques ultimately to provide the protein-nucleic acid molecular image that we seek. Our conviction that this structure analysis will be successful is supported by physical and chemical studies which indicate that there is no substantial change in the overall conformation of the gene 5 protein structure upon formation of complexes with oligonucleotides.

In the $P6_3$, $C222_1$, R32, and $P3_121$ crystals the number of gene 5 monomers/asymmetric unit is observed to be about 12. The repeated occurrence of this number of monomers (six dimers) as the asymmetric unit of the crystals suggests that it represents a specific aggregate of gene 5 protein that is formed upon addition of oligonucleotides. The fact that these aggregates crystallize requires that they be a homogeneous population of identically structured complexes and cannot be accounted for by an arbitrary or random aggregation process. Hence, they must represent some ordered mode of self assembly from the solution species.

To our knowledge, the complexes between the gene 5 DNA unwinding protein and oligonucleotides are among the first examples of crystalline protein·nucleic acid complexes available for study by x-ray crystallography. They are particularly significant in that they do diffract to sufficiently high resolution to permit a full three-dimensional structure analysis by single crystal methods. We expect this full analysis to contribute substantially to our understanding of the general stereochemistry of protein-nucleic acid interactions and to suggest some of the mechanisms by which these are mediated.

Acknowledgments—We thank Dr. J. H. Van Boom, Dr. Lloyd Williams, and Dr. Robert Ratliff for supplying deoxyoligonucleotides.

REFERENCES

1. McPherson, A., Jurnak, F. A., Wang, A. H. J., Molineux, I., and Rich, A. (1978) *J. Mol. Biol.* **134,** 379–400
2. Alberts, B., Frey, L., and Delius, H. (1972) *J. Mol. Biol.* **68,** 139–152
3. Oey, J. L., and Knippers, R. (1972) *J. Mol. Biol.* **68,** 125–138
4. Mazur, B. J., and Model, P. (1973) *J. Mol. Biol.* **78,** 285–300
5. Mazur, B. J., and Zinder, N. D. (1975) *Virology* **68,** 490–502
6. Nakashima, Y., Dunker, A. K., Marvin, D. A., and Konigsberg, W. (1974) *FEBS Lett.* **40,** 290–292
7. Cavalieri, S. J., Goldthwait, D. A., and Neet, K. E. (1976) *J. Mol. Biol.* **102,** 713–722
8. Day, L. A. (1973) *Biochemistry* **12,** 5329–5339
9. Pratt, D., Laws, P., and Griffith, J. (1974) *J. Mol. Biol.* **82,** 425–439
10. Anderson, R. A., Nakashima, Y., and Coleman, J. E. (1975) *Biochemistry* **14,** 907–917
11. Dunker, A. K., and Anderson, E. A. (1975) *Biochim. Biophys. Acta* **402,** 31–34
12. Henry, T. J., and Pratt, D. (1969) *Proc. Natl. Acad. Sci. U. S. A.* **62,** 800–807
13. Pratt, D., Tzagoloff, H., and Beaudoin, J. (1969) *Virology* **39,** 42–53
14. Coleman, J. E., Anderson, R. A., Ratcliffe, R. G., and Armitage, I. M. (1976) *Biochemistry* **15,** 5419–5430
15. Pretorius, H. T., Klein, M., and Day, L. A. (1975) *J. Biol. Chem.* **250,** 9262–9269
16. McPherson, A. (1976) *Methods Biochem. Anal.* **23,** 249–345
17. McPherson, A. (1976) *J. Biol. Chem.* **251,** 6300–6303
18. McPherson, A., Molineux, I., and Rich, A. (1976) *J. Mol. Biol.* **106,** 1077–1081
19. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497