Cocrystals of the DNA-binding domain of phage 434 repressor and a synthetic phase 434 operator
(protein–DNA cocrystals/x-ray crystallography/B-DNA/helical diffraction)

JOHN ANDERSON, MARK PtASHNE, AND STEPHEN C. HARRISON
Fairchild Biochemistry Building, Harvard University, Cambridge, MA 02138

Contributed by Mark Ptashne, November 7, 1983

ABSTRACT The amino-terminal domain of the phage 434 repressor forms cocrystals with a synthetic phase 434 operator. The cocrystals diffract to at least 4 Å, and x-ray crystallographic analysis of them is in progress. An analysis of the packing in the cocrystals shows that complexes consisting of dimers of amino-terminal domain bound specifically to operators are stacked end to end in long protein–DNA rods parallel to the unit cell body diagonals. The DNA in the complexes has 10.5 base pairs per turn and a rise per base of 3.26 Å—values consistent with B-form DNA—indicating that DNA is neither unwound nor overwound by bound repressor. The packing analysis suggests an approach that might facilitate the cocrystallization of other DNA-binding proteins with the DNA they recognize.

The crystal structures of three proteins that recognize specific DNA sequences are known (1–3). All three proteins regulate gene expression in Escherichia coli: λ repressor transcribes RNA polymerase, and the catabolite activator protein and λ repressor both repress and stimulate transcription (4–6). Each protein binds as a dimer to an ~2-fold rotationally symmetric DNA sequence. Monomers of the catabolite activator protein, λ repressor, and λ cro have a pair of protruding α helices that is imagined to make important DNA contacts. From protein to protein, this biblical substructure differs in amino acid sequence but is disposed nearly identically in space. Model building studies of the interaction of these proteins with DNA (2, 3, 7) suggest that one of the helices penetrates the major groove in an operator half-site, making specific contacts, and the other lies across the major groove, making nonspecific contacts with the phosphate backbone. The 2-fold-related monomer makes similar contacts in the other half-site. Amino acid homologies with the region containing the biblical substructure have been found in many other DNA-binding proteins (8–10), suggesting that those proteins recognize DNA using a similar pair of α helices. No cocrystal of any of these DNA-binding proteins with a complete operator site, of sufficient quality for high resolution crystallographic analysis, has been reported previously, although cocrystals of λ cro with partial operator sites have been reported recently (11).

The E. coli bacteriophages 434 and λ are closely related, and 434 repressor is similar in most respects to λ repressor (Fig. 1). Like λ repressor (12, 13), 434 repressor has two domains that are separable by proteolytic cleavage (see Results). In each repressor, the amino-terminal domain binds DNA (ref. 13; R. P. Wharton, personal communication), and the carboxyl-terminal domain promotes dimer formation (ref. 12; unpublished data). Amino acid sequence homologies suggest that 434 repressor recognizes DNA using a biblical substructure similar to that found in λ repressor, λ cro, and catabolite activator protein (9). Both λ and 434 repressors bind to tripartite operators, OR and OL, on the phage DNA (4, 5). Phage 434 operator sites are 14 base pairs long and have the consensus sequence 5′-A-C-A-N-N-N-N-N-N-T-T-G-T-3′ (ref. 5; unpublished data). Phage 434 repressor, like λ repressor (4, 5), can both repress and stimulate transcription by RNA polymerase (unpublished data). The sequences of the 434 and λ operators are quite different, however (5), as are the protein sequences (ref. 14; unpublished data).

We report here the isolation of a fragment of 434 repressor comprising its amino-terminal domain and its cocrystallization with a complete 14-base-pair 434 operator, and we describe the packing in these cocrystals. From the packing analysis, we conclude that the DNA in the complex is B-form, with 10.5 base pairs per turn. Finally, we outline a strategy, suggested by the packing scheme, that might facilitate cocrystallization of other repressor–DNA complexes.

MATERIALS AND METHODS

Phage 434 repressor was produced from cells bearing pJA41. To make this plasmid, a λ 401-base-pair Hpa II fragment carrying the 434 cl gene and the −10 region of the E. coli lacUV5 promoter was isolated from pRP42 (15) and inserted into the Cla I site of pEA300, reconstituting the tac promoter (16). The E. coli lacI2 strain RB791 (17) carrying pJA41 produces ~10% of the cellular protein as 434 repressor when fully induced by isopropyl β-D-thiogalactoside as judged by NaDodSO4/polyacrylamide gel electrophoresis of whole cell lysates (18).

RB791/pJA41 was grown in LB with ampicillin at 50 μg/ml (Sigma) at 37°C to an OD600 of 1. Isopropyl β-D-thiogalactoside (Sigma) was added to a final concentration of 0.1 mM, and growth was continued for 8 hr. The cells were harvested, and 434 repressor was purified by polyethyleneimine precipitation, ion-exchange chromatography on carboxymethyl-Sephadex (Pharmacia), and ammonium sulfate reverse extraction. Details of the repressor purification will be described elsewhere.

Phage 434 repressor amino-terminal domain (434R1-40) was prepared as described in the legend to Fig. 2.

The synthetic 434 operator, 5′-HO-d(A-C-A-T-A-T-A-T-A-T-G-T)-OH-3′, used in cocrystallization attempts was purchased from P-L Biochemicals. It was supplied as a solid, which was dissolved at 10 mg/ml in 0.01 M Tris-HCl, pH 7.4/0.001 M EDTA/0.01 M NaCl.

Cocrystals were grown by vapor diffusion in hanging drops (19) at 4°C over wells containing 1.3 M (NH4)2SO4 and 0.005 M Na2HPO4 at pH 4.7. Each drop initially contained 434R1-40 at 4.5 mg/ml and operator at 2.3 mg/ml (molar ratio protein monomer/DNA, 2:1). X-ray diffraction data were collected by oscillation or precession photography on CEA Reflex-25 film (CEA America) using an Elliot GX-6 or GX-13 rotating anode x-ray generator and a precession/oscillation camera (Charles Supper).

Determination of amino-terminal amino acid sequences by...
automated Edman degradation and carboxyl-terminal amino acid sequences by carboxypeptidase digestion and of amino acid compositions by automated amino acid analysis was carried out as described (20, 21).

The concentrations of 434R169 and of operator were calculated from the optical densities at 280 and 260 nm, respectively. The extinction coefficient of 434R169, calculated from its amino acid composition (unpublished data), is 0.73 (mg/ml)⁻¹, and that of the operator, supplied by the manufacturer, is 19.0 (mg/ml)⁻¹.

Restriction enzymes and T4 DNA ligase were purchased from New England BioLabs and were used as suggested by the manufacturer.

**RESULTS**

Preparation of 434R169 and Its Cocrystallization with a Complete 434 Operator. Papain digestion of purified 434 repressor produces two polypeptides with approximate molecular weights of 15,000 and 10,000 (Fig. 2a). Analysis of the primary structure of these fragments, using automated Edman degradation, carboxypeptidase digestion, and amino acid composition analysis, shows that the larger fragment contains residues 73–209 and that the smaller fragment contains the first 69 residues of the intact repressor (data not shown). The relative stability of the fragments, even though there are potential papain cleavage sites (23) within both, indicates that they are comprised of folded domains.

The carboxyl-terminal fragment flows unimpeded through an SP-Sephadex column in 0.05 M NaCl at pH 7.9, but the amino-terminal fragment binds tightly to the resin. A linear gradient of increasing NaCl elutes the amino-terminal fragment (434R169) as a homogeneous preparation (Fig. 2b).

Purified 434R169 is a monomer at concentrations (40 μM) where intact repressor is a dimer (unpublished data). It binds 434 operators specifically in DNase protection experiments (R. P. Wharton, personal communication).

![Fig. 1. A repressor dimer bound to its operator site. The operator site is partially 2-fold rotationally symmetric (indicated by arrows). A repressor monomer consists of two domains separated by a protease-sensitive linker. The carboxyl-terminal domains (C) of repressor monomers contact each other to form dimers, the physiologically active, DNA-binding species. Each amino-terminal domain (N) of a dimer makes specific DNA contacts in the major groove of one half of the operator site. The general features of this model hold for λ, P22, and 434 repressors, but the amino acid sequences of the proteins, and the lengths and DNA sequences of the operator sites are different. The λ repressor, unlike the P22 and 434 repressors, also has an amino-terminal arm that wraps around the DNA.](image)

**FIG. 2.** Preparation of 434R169. (a) NaDodSO4/polyacrylamide gel electrophoresis of the products of papain digestion of 434 repressor. Two polypeptides with molecular weights of about 15,000 (C) and 10,000 (N) remain even after 26 h of incubation of repressor (Rep) with papain at room temperature. Phage 434 repressor at 5 mg/ml in 0.005 M Na2HPO4/0.005 M NaH2PO4/0.001 M NaNO3/0.0001 M EDTA/0.0001 M dithiothreitol/5% glycerol/0.2 M NaCl was digested with activated papain (Worthington) at an enzyme/substrate ratio of 1:3000. Papain was inactivated by the addition of 2-bromo-2'hydroxy-5'-nitroacetamide (Koshland's Reagent III, Aldrich) as described (22). The digestion mixture was dialyzed to 0.01 M Tris-HCl, pH 7.9/0.0001 M EDTA/0.001 M NaNO3/0.0001 M dithiothreitol/5% glycerol/0.05 M NaCl and loaded onto a 200 ml SP-Sephadex C-25 column (Pharmacia) equilibrated in the same buffer. The carboxyl-terminal fragment (C) flowed through the column, but the amino-terminal fragment (N) did not. The amino-terminal fragment (434R169) eluted at about 0.175 M NaCl in a linear 0.05–0.25 M NaCl gradient of total volume 1440 ml. Fractions containing 434R169 were pooled, dialyzed to 0.01 M Tris-HCl, pH 7.9/0.001 M NaNO3, and concentrated by ultrafiltration under N2 pressure through a YM5 membrane in an Amicon ultrafiltration device to a final concentration of 14 mg/ml. (b) Gel electrophoresis of purified 434R169.

We have obtained cocrystals of 434R169 and a synthetic oligodeoxy nucleotide with the 2-fold symmetric sequence


This sequence differs by only one base from that of O12, one of the operator sites in the phase 434 left operator (unpublished data), in which A replaces T at position 7. Analysis of dissolved cocrystals on polyacrylamide gels indicates that they contain a 2:1 ratio (mol/mol) of protein monomers to operator. The cocrystals belong to the body-centered tetragonal space group I422. The unit cell dimension are a = b = 166 Å, c = 139 Å. The cocrystals are well ordered and diffract to about 4 Å in all directions and 3 Å in certain directions (see below).

**Overview of the Packing in the Cocrystals.** The fundamental element of packing in the I422 cocrystals is a complex of the 14-base-pair operator and a dimer of 434R169. We believe that the 2-fold axis of the protein dimer is coincident with the 2-fold axis of the operator (see Discussion). These complexes are stacked end to end in rods that run parallel to, but not coincident with, the body diagonals of a unit cell (Fig. 3) and extend unbroken over a considerable distance in the crystal. The complexes are stacked so that the register of the DNA helix is maintained across the junctions between complexes. In the length of the body diagonal, 274 Å, there are six complexes containing 84 base pairs of the pseudo-continuous DNA helix with 10.5 base pairs per turn and an average rise per base pair of 3.26 Å. Each 14-base-pair operator contains one and a third turns of the DNA helix; three stacked operators therefore contain four turns. The dyad of each complex is a third of the way farther around the rod
axis than the last; thus, the position around the rod of every third complex is identical.

**Packing Analysis.** The diffraction patterns of the I422 cocryystals have a number of striking features from which we derived the packing scheme. In this section, we take each of these features in turn and show how it is accounted for by our packing model, and then we discuss the packing in somewhat more detail.

In screenless precession photographs of the I422 cocrystals, there are strong spikes of intensity in the zero level and diffuse diffraction at about 3.4 Å (Fig. 4 a–c). All of the diffracting material in the cocrystals is associated with sets of planes formed by the protein–DNA rods, and reflections along lines normal to these planes should be particularly intense. The inset of each diffraction photograph in Fig. 4 a–c shows a computer-generated projection through our packing model in the direction in which the photograph was taken, demonstrating that the spikes are normal to the planes. The faintly visible 3.4 Å diffraction is due to the base-pair repeat of the DNA in the rods.

If the crystal is tilted so that the 3.4 Å diffraction is in reflecting position, the dominant feature of oscillation photographs is a series of very sharp layer lines, sampled at points of the reciprocal lattice, arranged in the pattern characteristic of diffraction from a helix with three units per turn, a pitch of about 135 Å, and a rise per helical repeat of about 45 Å (Fig. 4e). The strongest line is at 3.26 Å. The protein–DNA rods in our packing model, which form noncrystallographic 3-fold screw axes with a pitch of 137 Å and a rise per repeat unit of 45.7 Å, are the source of this helical diffraction pattern. The helical diffraction pattern is strong at this resolution because it is sampling the very strong underlying transform of the DNA base-pair repeat, and the strongest layer line corresponds to that repeat distance. The sharpness of the layer lines indicates that the coherence of the stacking is maintained over a distance along the rods much greater than the dimension of one unit cell. The DNA diffraction appears diffuse in the precession pictures in Fig. 4 a–c simply because it is not in the optimum reflecting position.

Diffraction extends to about 3 Å resolution in the direction of the protein–DNA rods. In other directions, it extends only to about 4 Å. The packing is looser in directions perpendicular to a rod than parallel to a rod, allowing flexibility that could limit order in those directions. This idea is supported by the observation that cooling the cocrystals during data collection, or using high intensity synchrotron radiation (P. J. Bjorkman, personal communication), enhances the reciprocal lattice sampling of the rod-associated layer lines, but does not extend the 4 Å data between rods at all.

Our packing model as described so far fully accounts for the most striking features of the diffraction patterns of the I422 cocrystals. We can define the packing more precisely by considering the position of complexes along the rods with respect to the unit cell. There are two types of dyads perpendicular to the rod axis. One is the 2-fold axis of the complex; the other is the 2-fold axis between complexes (Fig. 3b). The positions of complexes along the rods with respect to the unit cell are fixed by requiring that these dyads be coincident with crystallographic dyads that are parallel to [110] and intersect [111] at z = 0 and z = 1/4. There are four possible packings, either of which the two rod dyads may coincide with either crystallographic dyad in either of two orientations 180° apart, but this ambiguity can be resolved only by solving the structure.

The part of a rod extending from the corner lattice point to the center lattice point of a unit cell is a crystallographic repeat. It contains exactly three dimer–operator complexes, and it must also contain an integral number of turns of the pseudo-continuous DNA helix. This requirement would be satisfied by having 3, 4, or 5 turns of DNA, giving 14, 10.5, or 8.4 base pairs per turn, respectively. Although our data do not rule out 3 or 5 turns, we believe that these values are unlikely for two reasons. First, the average rotations per base pair for DNA helices with 8.4 and 14 bases per turn, 43° and 26°, are out of the range generally observed (30°–36°; ref. 25a). Second, experiments have shown that a repressor, a protein nearly identical in structure and function to 434 repressor, does not significantly change the helical repeat of DNA when it binds (26). A helical repeat of 10.5 bases per turn would give an average rotation per base of 34°, well within the observed range. Moreover, it is close to the value observed for most sequences in solution (27, 28).

The asymmetric unit of the I422 cocrystals contains one and a half dimer–operator complexes, implying that 80% of the unit cell volume is occupied by solvent.

The packing we have described is a tetragonal version of the cubic close packing of rods discussed by O'Keefe and Andersson (29).

We have obtained two other, less well ordered cocrystal forms: P41222 cocrystals grow as stacks of plates; P41 cocrystals grow as truncated bipyramids with faint layering visible parallel to the base. Diffraction photographs of the [001] zone (normal to the plate face for P41222, normal to the pyramidal base for P41) have reasonable order, but in the perpendicular direction both cocrystal types are badly disordered. The orientation of diffuse diffraction at about 3.4 Å in these diffraction patterns indicates that in both cocrystals the DNA is packed with its helix axis parallel to the crystallographic axes a and b; both of which are 136 Å long—close to the distance from corner to center in the I422 unit cell. If the P41 cocrystals are tilted so that the apparently diffuse diffraction at 3.4 Å is in reflecting position, a helical diffraction pattern identical to that in the I422 cocrystals is seen (not shown). Thus, the less well ordered cocrystals contain the...
FIG. 4. Diffraction photographs of the 1422 cocrystals. Crystal to film distance was 10 cm. CuKα radiation, λ = 1.54 Å. Temperature was 22°C in a-c and 4°C in d and e. (a-c. Insets) Projection of the packing in the cocrystal in the direction of the photograph (protein is omitted for clarity), showing the relationship of the zero level spikes of intensity and 3.4 Å diffuse diffraction (arrowheads) to the planes in the cocrystal. (a) 3° screenless precession photograph taken along [100]. Spikes normal to protein-DNA planes are at 50.1° to horizontal. (b) 3° screenless precession taken along [110]. Spikes are both horizontal and inclined at 59.4°. (c) 2° screenless precession taken along [001]. Spikes are horizontal and vertical. The a and b axes are inclined at 45°. (d) 2° oscillation photograph (oscillation axis 011) taken with the crystal tilted so that the 3.4 Å diffuse diffraction is in optimum reflecting position. Brackets embrace the helical diffraction pattern arising from the 3-fold screw symmetry around protein-DNA rods in the cocrystals. (e) Enlargement of the bracketed region in d. Dimensions are given in Å⁻¹. Sharp layer lines, sampled by the reciprocal lattice, constitute a helical diffraction pattern (24). Meridional reflections occur on every third layer line, indicating that the helix has three repeat units per turn. The separation between successive layer lines corresponds to a helix pitch of 135 Å, and the separation between meridional reflections corresponds to a rise per repeat unit of 45 Å. The upper, most intense layer line is at 3.26 Å.

same rod-like packing unit as the 1422 cocrystals, implying that the rods must be very stable. (The disorder in the P4₁ and P4₁2₁2 cocrystals probably results from slippage between parallel layers of rods. Such slippage is not possible in the 1422 cocrystals because of the highly interwoven nature of the packing.) The identity of the rod structure in both types of cocrystal (layered and interwoven) suggests that it is not the result of constraints imposed by the packing in either type, but that it is a stable structure in itself. Our data do not exclude the possibility that the helix axis of the DNA in a complex is displaced from the rod axis along the dyad of the complex. It seems to us unlikely, however, that such a crankshaft-like rod could pack unchanged in two so different schemes.

DISCUSSION

In the 1422 cocrystals we have described, a dimer of 434R₁₄₋₄₀₉ is bound to each 14-base-pair operator. The space group symmetry requires that the dyad of a dimer of 434R₁₄₋₄₀₉ be superposed on a dyad associated with the DNA rod. We believe that the 434R₁₄₋₄₀₉ dimer is bound with its dyad coincident with the operator dyad in much the same way as λ cro and λ repressor DNA-binding domain are imagined to bind to their DNA binding sites (1, 3). [Strictly speaking, we cannot exclude the possibility that each dimer is bound with its dyad coincident with the dyad between operators (see Fig. 3b).] We expect that solution of the cocrystal structure will confirm this aspect of our model and will reveal the specific interactions between the protein and the DNA.

Our work with these cocrystals has already produced an important result: the DNA bound by 434R₁₄₋₄₀₉ in our cocrystals has 10.5 base pairs per turn and 3.26 Å per base, indicating that 434R₁₄₋₄₀₉ binds to B-form DNA. This result is not unexpected because the helical repeat found for most DNA sequences in solution is about 10.5 base pairs per turn (27, 28), and biochemical and biophysical experiments have shown that there is no drastic change in DNA conformation on binding by λ repressor (26), lac repressor (30), and catabolite activator protein (31).

We have had difficulty in the past growing repressor-DNA cocrystals of sufficient quality for high resolution x-ray diffraction analysis. We believe that the formation of our cocrystals was facilitated by the length of the operator fragment. Each 14-base-pair operator contains exactly one and a third turns of DNA at 10.5 bases per turn. Three operators, stacked as we have described, therefore contain exactly four turns of pseudo-continuous DNA, and the structure of the protein-DNA rod repeats after every third complex. We
suggest the following approach to the cocrysalization with DNA of other proteins that bind DNA without altering its helical repeat. Pieces of DNA should be used that are of sufficient length to carry a functional binding site and that are 7n bases long (in which n is an integer greater than 1). (Because most binding sites are no more than 21 bases long, the usual value for n will be 2 or 3.) Each fragment will contain 2n/3 turns of DNA at 10.5 base pairs per turn. If n is a multiple of three, each fragment will be a repeat in itself. If n is not a multiple of three, rods formed from such fragments will have 3-fold screw symmetry (like the ones in the I422 crystals), and a repeat will occur after three complexes. In either case, the tendency of the bases at the ends of the DNA fragments to stack as we have described in this paper should promote the formation of an ordered substructure with the protein bound in a physiologically significant way.

We thank Bruce Driver for assistance with computer graphics and Joette Stromberg-Anderson for illustrations. We also thank Roger Brent and Robin Wharton for useful comments on the manuscript. This work was supported by Grants GM29109 (to M.P.) and CA13202 (to S.C.H.) from the National Institutes of Health.