Structure of the Lambda Complex at 2.5 Å Resolution: Details of the Repressor-Operator Interactions

STEVEN R. JORDAN AND CARL O. PABO

The crystal structure of a complex containing the DNA-binding domain of lambda repressor and a lambda operator site was determined at 2.5 Å resolution and refined to a crystallographic *R* factor of 24.2 percent. The complex is stabilized by an extensive network of hydrogen bonds between the protein and the sugar-phosphate backbone. Several side chains form hydrogen bonds with sites in the major groove, and hydrophobic contacts also contribute to the specificity of binding. The overall arrangement of the complex is quite similar to that predicted from earlier modeling studies, which fit the protein dimer against linear B-form DNA. However, the cocrystal structure reveals important side chain–side chain interactions that were not predicted from the modeling or from previous genetic and biochemical studies.

CRYSTALLOGRAPHIC STUDIES OF DNA-BINDING PROTEINS (*1–4*) and of protein-DNA complexes (*5, 6*) are beginning to give structural information about site-specific recognition and gene regulation, but many important problems remain. Even in cases where we can assign a binding protein to a known structural class and determine that it contains a "helix-turn-helix" motif (*7*) or a "zinc finger" motif (*8*), we still cannot reliably predict how the protein contacts its binding site or which residues should be altered to change the specificity. Detailed information about several protein-DNA complexes will be needed to determine whether any general principles emerge.

The repressor from bacteriophage lambda provides a useful system for the study of protein-DNA interactions. Repressor regulates gene expression by binding, with different affinities, to six different sites in the phage DNA (*9*). Each site contains 17 base pairs and is approximately symmetric about the central base pair. The amino-terminal domain of repressor, residues 1 to 92, is the DNA-binding domain and makes all of the contacts required for positive and negative control of transcription (*10*). The crystal structure of this domain has been reported, and a model for the repressor-operator interactions has been proposed (*3*). Lambda repressor also has been the subject of extensive genetic and biochemical analysis, and there have been several studies of mutations that affect recognition and regulation (*11–14*).

Although these genetic and biochemical studies have supported the basic proposals that came from model building, cocrystals were needed to test these models and to understand details of the repressor-operator interactions. After testing a series of operator fragments of different length, we crystallized the NH_2-terminal domain of lambda repressor with a 20-bp fragment (*15*) (Fig. 1) that contains the operator site O_L1. Structural analysis of these crystals reveals many details, including some unexpected side chain–side chain interactions, that have important implications for understanding site-specific recognition and gene regulation.

Structure determination. Crystals of the repressor-operator complex were grown by vapor diffusion against 20 percent PEG 400 (*15*). The complex crystallized in space group $P2_1$ with unit cell lengths a = 37.22 Å, b = 68.72 Å, and c = 57.03 Å and with a β angle of 92.2°. The volume of the unit cell suggested that there was one complex in the asymmetric unit. Transferring macroscopic seeds into fresh crystallization drops gave large crystals that were used for data collection, and these crystals diffracted to at least 2.25 Å in all directions.

Precession photographs indicated that the crystals contained B-DNA oriented along the diagonal of the ac plane (*15*). Using this restraint and packing considerations, we located the proposed model of the complex (*3*) in a low-resolution *R*-factor search, with 5 Å data collected on a Nicolet P3/F diffractometer. This model was partially refined with CORELS (*16*), and it served as a starting point for rebuilding the complex at high resolution.

The structure that we now report is based on two data sets collected on the Xentronics area detector at Genex Corporation (Gaithersburg, Maryland). The native data were collected to 2.5 Å resolution (Table 1), and data from a $Pb(NO_3)_2$ derivative were collected to 3.1 Å. Metal sites were located with the use of the CORELS model to phase a difference map, but the sites were readily apparent from the difference Patterson maps. The sites were refined with the Munich PROTEIN system (*17*), and a SIR map, with a

Table 1. Statistics for the native data as a function of resolution shell. Observed reflections had $I > 2\sigma$. The R_{sym} is based on 22,860 observations of 9,177 reflections.

$$R_{sym} = \sum_i \sum_j |I_{ij} - \bar{I}_i| / \sum_i \sum_j I_{ij}$$

Resolution (Å)	Number observed	Number possible	Percentage observed	R_{sym}
4.54	1684	1703	98.9	0.038
3.61	1645	1677	98.1	0.039
3.15	1572	1648	95.4	0.042
2.86	1512	1663	90.9	0.056
2.66	1421	1652	86.0	0.071
2.50	1343	1656	81.1	0.082
Infinity to 2.50	9177	9999	91.8	0.043

S. R. Jordan is in the Department of Biophysics, and C. O. Pabo is in the Department of Molecular Biology and Genetics and the Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

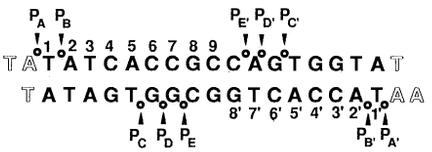


Fig. 1. Operator site used for cocrystallization. Sequence of the 17-bp site O_L1 is shown in bold letters. The left half of the operator site, with base pairs numbered 1 to 9, matches the consensus sequence for the operator half-sites (*9*). The approximate twofold axis goes through base pair 9, and base pairs in the nonconsensus half of O_L1 are numbered 1′ to 8′. Circles mark phosphates that ethylation interference experiments had implicated as contacts (*31*). These are labeled P_A to P_E in the consensus half-site and $P_{A'}$ to $P_{E'}$ in the nonconsensus half.

mean figure of merit of 0.57, was calculated at 3.1 Å resolution. Although this map was not adequate for rebuilding the complex, we followed Wang's procedure of iterative phase improvement with automatic masking (*18*), and the map was dramatically improved.

This electron density map indicated that the CORELS model was basically correct. It confirmed that there was one complex in the asymmetric unit and that the DNA duplexes from neighboring unit cells were stacked to form a continuous, but somewhat irregular, helix that runs through the crystal. Coordinates from the CORELS model were used as a starting point for rebuilding the structure, but the electron density map had not been biased in any way by the CORELS model, as the masks used in Wang's program are not dependent on any molecular model or previously defined envelope. The central 17 bp of the DNA and the first four helices of the protein were clearly visible in the electron density map, and only minor adjustments were made in these regions before starting refinement. Portions of the complex that were not clear in this map and could not be rebuilt reliably were not included in the early refinement cycles. Specifically, we omitted the NH_2-terminal arm (residues 1 to 8), the loop between helix 1 and helix 2 (residues 24 to 30), and helix 5 and the preceding loop (residues 71 to 92). We also omitted bases flanking the O_L1 site, removing two bases from the 5′ end of each DNA strand and one from the 3′ end (Fig. 1).

The model was refined with the use of the Hendrickson and Konnert conjugent gradient program (*19*) that we modified to accept nucleic acid restraints based on a version of NUCLIN developed by Quigley and Westhof (*20*). The first cycles of refinement with the truncated model were based on data between 10.0 and 3.25 Å resolution. The resolution was gradually extended, and portions of the model that had been omitted in the first cycles were rebuilt as they became clear in $2|F_o| - |F_c|$ difference maps. Since the DNA duplex is approximately symmetric, it was difficult to choose the correct orientation from the first electron density maps. However, as refinement proceeded, it appeared that one of the two orientations was more consistent with the $2|F_o| - |F_c|$ maps, and the duplex was rotated accordingly. The orientation was later confirmed by growing cocrystals that contained 5-bromocytosine at base pair 4, and a difference map indicated that at least 90 percent of the duplexes have this orientation (*21*). In later stages, cycles of refinement with XPLOR (*22*), which combines crystallographic restraints, energy terms, and molecular dynamics, were interspersed with Hendrickson-Konnert refinement. Eventually, the entire complex was rebuilt from difference maps, although the density for residues 1 to 5 in the nonconsensus half-site is quite weak, and the position of these residues must be considered tentative. The final refinement cycle was obtained with the Hendrickson-Konnert program, and this model, with a single isotropic temperature factor, has an *R* factor of 24.2 percent for data from 6.0 to 2.5 Å (*23*). Difference maps indicate plausible locations for several water molecules, but these are not included in the current structure.

Overall structure of the complex. The overall structure of the repressor-operator complex (Fig. 2) shows that the protein dimer binds by fitting one helix from each subunit into the major groove of the operator DNA. The complex is roughly symmetric, and the twofold axis of the protein dimer coincides with the approximate twofold axis of the operator site. This basic arrangement was predicted after the structure of the lambda repressor was determined (*3*), and a similar arrangement was observed in cocrystals that contain the 434 repressor (*6*).

As observed in the protein crystals (*3*), the repressor monomer contains five α helices. The first four helices form a compact globular domain, with helices 1, 2, and 3 on the face that is closer to the DNA. Helices 2 and 3 form the conserved "helix-turn-helix" unit that has been observed in a number of DNA-binding proteins (*1–4, 6, 7*). Helix 5 extends out from this globular domain to contact helix 5′ of the other subunit and form the dimer interface. The structure of the protein dimer as observed in the complex is very similar to that reported for the uncomplexed repressor (*3*). The largest differences are in the loop regions, but the significance of these differences is not clear because the structure of the protein has only been determined to 3.2 Å resolution and has not been refined.

As previously predicted, residues in helices 2 and 3 (the "helix-turn-helix" unit) make many of the contacts with the DNA, but the cocrystal structure reveals that residues in the COOH-terminal part of helix 1 and residues in the loop following helix 3 also make important contacts. As predicted from biochemical studies and modeling (*3, 24, 25*), the NH_2-terminal arm wraps around the DNA and makes contacts in the major groove on the back side of the double helix. The operator DNA is fundamentally B-form DNA, with no striking kinks such as those observed in the Eco RI cocrystals (*5*). The largest deviations from linear B-DNA occur at the ends of the operator site, which bend slightly toward repressor (Fig. 2).

Contacts with bases in the major groove. Repressor contacts sites in the major groove and phosphate oxygens of the ribose phosphate backbone. Although the set of contacts must be responsible for specific recognition, we begin by discussing contacts with the edges of the base pairs. Critical hydrogen bonds are made by Gln^{44} and Ser^{45}, which are at the NH_2-terminal end of helix 3; by Asn^{55}, which is in a loop just after helix 3; and by Lys^4, which is in the NH_2-terminal arm (Fig. 3). Since most of the contacts are the same in the two halves of the operator site, we discuss contacts with

Table 2. Local helical parameters of O_L1. Helical twists are described with respect to a single helical axis. Propeller twists are described with respect to a line connecting the C6 of the pyrimidine and the C8 of the purine.

Position	Base pair	Twist (deg)		Position	Base pair	Twist (deg)	
		Helical	Propeller			Helical	Propeller
1	T·A	.....	11.9	9	C·G	.....	5.4*
		41.4				27.3	
2	A·T	.....	18.0	8′	C·G	.....	21.2
		29.6				47.0	
3	T·A	.....	13.7	7′	A·T	.....	9.1
		37.8				34.2	
4	C·G	.....	12.6	6′	G·C	.....	11.8
		32.4				33.2	
5	A·T	.....	4.7	5′	T·A	.....	13.6
		33.4				31.0	
6	C·G	.....	9.4	4′	G·C	.....	13.6
		21.4				33.9	
7	C·G	.....	2.9	3′	G·C	.....	12.1
		36.4				33.3	
8	G·C	.....	18.7	2′	T·A	.....	11.9
		39.4				39.1	
9	C·G	.....	5.4	1′	A·T	.....	6.5

*For clarity, the central base pair is repeated at the start of the second column.

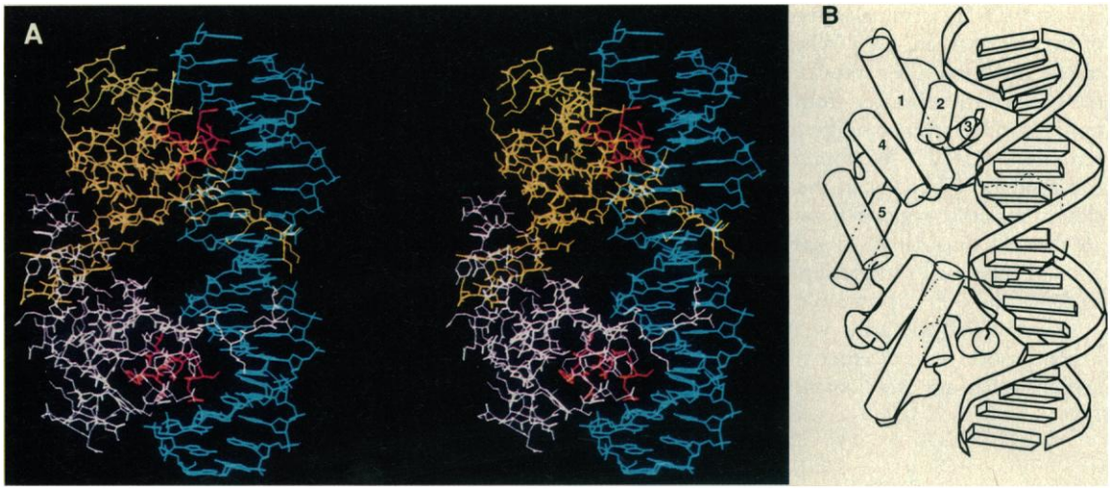


Fig. 2. (**A**) Stereo photograph of the repressor-operator complex. The DNA is dark blue, the repressor monomer bound to the consensus half-site is yellow, and the monomer bound to the nonconsensus half-site is purple. In each chain, helix 3 (residues 44 to 52) is red. (**B**) Sketch of the repressor-operator complex in the same orientation as (A). Helices in one monomer are numbered 1 to 5. The terminal base on the 5′ end of each strand has been omitted.

respect to the consensus half-site and note significant differences when they occur.

As predicted, Gln^{44}, which is the first residue in helix 3, forms two hydrogen bonds with the adenine of base pair 2 (Fig. 3). The amide $-NH_2$ of the glutamine side chain donates a hydrogen bond to N7 of adenine and the amide =O accepts a hydrogen bond from the N6. This type of bidentate interaction was proposed by Seeman, Rosenberg, and Rich (*26*), was observed in the 434 cocrystals (*6*), and is predicted to occur on the λ cro complex (*1*). However, the λ cocrystals reveal an interesting and unexpected aspect of this interaction: The contact made by Gln^{44} is stabilized by a hydrogen bond from the side chain of Gln^{33} (the first residue in helix 2); Gln^{33} also hydrogen bonds to the oxygen of phosphate P_B (Fig. 3). This gives an extended hydrogen bonding network with two protein side chains connecting a phosphate with a base in the major groove. A similar extended interaction involving a glutamine at the start of helix 2 and a glutamine at the start of helix 3 has been observed in a high-resolution study of the 434 cocrystals (*27*). These extended interactions help to explain why the conserved helix-turn-helix unit frequently begins with a glutamine (*7*, figure 12).

As predicted, Ser^{45} (which is the second residue in helix 3) hydrogen bonds to the N7 of guanine 4, and may also have a weaker interaction with the O6 of this guanine. The last sequence-specific hydrogen-bonding interaction from this part of the protein involves Asn^{55}, which is in a loop three residues after the end of helix 3. As predicted, this asparagine hydrogen bonds to the guanine in base pair 6. However, hydrogen bonds with N7 and O6 were predicted, and the cocrystal structure shows that asparagine only hydrogen bonds to N7. In the consensus half-site, Lys^4 makes an interesting and totally unexpected contact with this guanine. The arm reaches toward the center of the operator site, but Lys^4 extends back and binds in a pocket formed by O6 of guanine 6 and the amide =O of Asn^{55} (Fig. 3). As previously noted, the density for the arm is not well defined in the other half-site.

The cocrystal structure shows that hydrophobic interactions also contribute to the specificity of the repressor-operator interaction. The clearest examples involve the thymine methyl groups in base pairs 3 and 5 of the consensus half-site. At base pair 3, a hydrophobic pocket is formed by the β carbon of Ala^{49} and the terminal carbon atom of Ile^{54} (in the loop just after helix 3); Gly^{48} is also very close to this thymine methyl group and may help to exclude water from the site. At base pair 5, the closest neighbors of the thymine methyl group include Gly^{46} and the β carbon of Ser^{45}. [Ser^{45} hydrogen bonds to guanine 4, but the helical twist of the DNA brings the β carbon very close to the methyl group of thymine 5.] The methyl group of Ala^{49} is slightly farther away but it probably contributes to the hydrophobicity of this region. At base pair 1, there seem to be weaker hydrophobic contacts with the methyl group of the thymine. It is not buried as completely as the methyl groups at base pairs 3 and 5, but the γ carbon of Gln^{44} and the methylene carbons of Glu^{34} are near this thymine.

Contacts with the sugar-phosphate backbone. The repressor makes hydrogen bonds with oxygens on each of the ten phosphates marked in Fig. 1. These contacted phosphates are labeled P_A to P_E in the consensus half-site and $P_{A'}$ to $P_{E'}$ in the nonconsensus half-site. The sheer number of contacts suggests that these interactions provide a significant portion of the overall binding energy. Many of these backbone contacts are made by residues that flank the helix 2–helix 3 recognition unit, and we presume that these backbone contacts help to orient and position the residues that contact the edges of the base pairs (Fig. 4).

The network of hydrogen bonds that repressor makes with the outer phosphates (P_A and P_B on the consensus half-site, $P_{A'}$ and $P_{B'}$ on the nonconsensus half) is particularly impressive. Since these backbone contacts are quite symmetric, at most positions we only describe contacts in the consensus half-site. Helix 1 runs along the "outside edge" of the sugar-phosphate backbone, and the Lys^{19} and Tyr^{22} side chains hydrogen bond to the oxygens that are further from the major groove. Residues in helices 2 and 3 contact oxygens on the "inside edge" of the DNA backbone. The peptide NH of Gln^{33} contacts P_A, and the dipole moment of the α helix (*28*) also contributes to this interaction. The side chain amide of Gln^{33} (the first residue in helix 2) and side chain of Asn^{52} (the last residue in helix 3) contact P_B. Although it is not close enough to form a strong hydrogen bond, the $-NH_3^+$ of Lys^{26} may also interact with phosphate P_A.

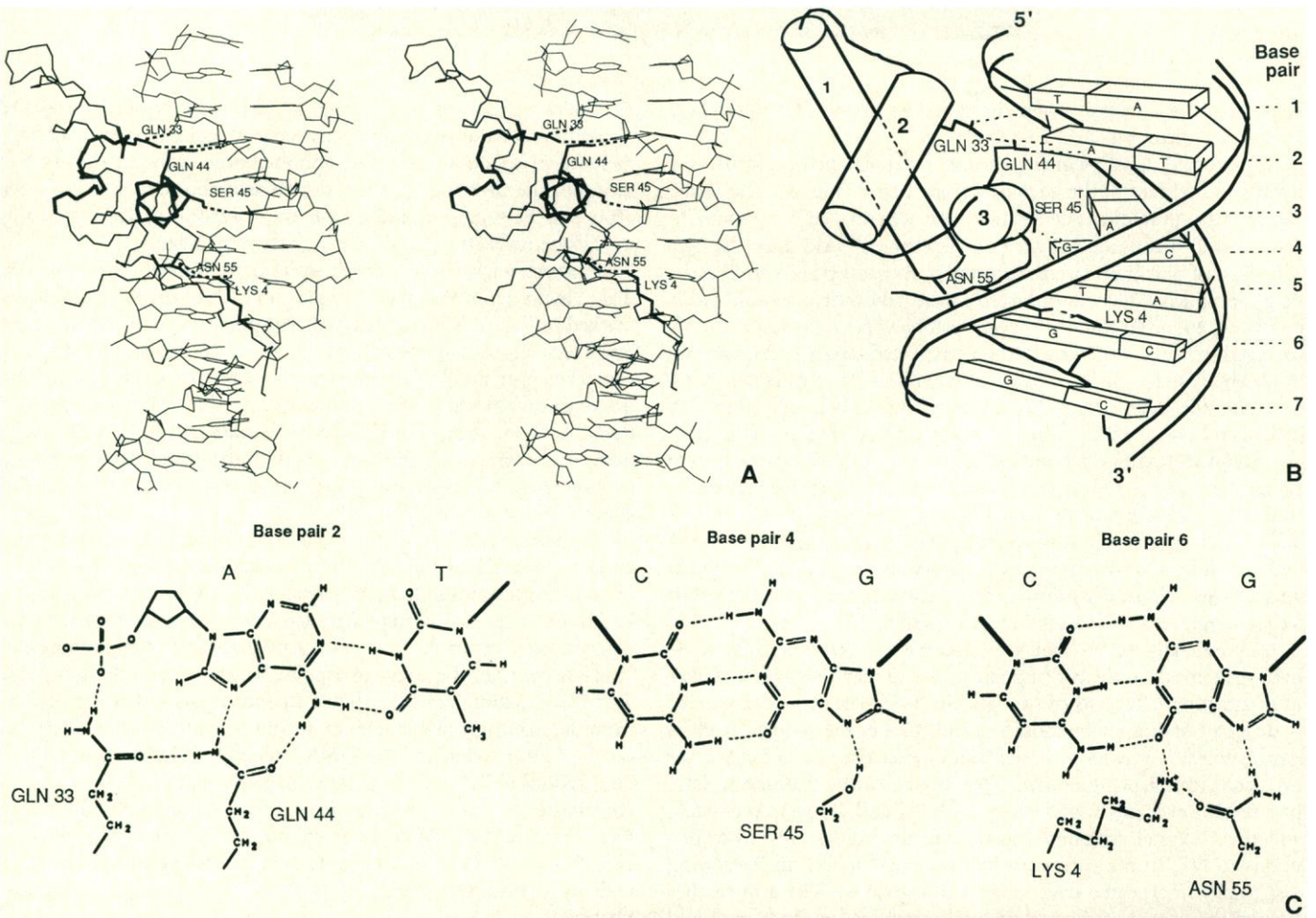
Important contacts are also made with phosphates P_C, P_D, and P_E, although the network of hydrogen-bonding contacts is less extensive. The peptide -NH of Gly^{43} (which is in the loop between helix 2 and helix 3) contacts the "inner" oxygen of P_C, but the protein does not form a hydrogen bond with the outer phosphate. The Asn^{61} side chain hydrogen bonds to P_D and may form a bifurcated hydrogen bond with the two oxygens. The hydrophobic side chain of Met^{42} contacts the DNA backbone between P_C and P_D and appears to interact with the C3′, C4′, and C5′ atoms on the edge of the sugar ring. The contacts with P_E and $P_{E'}$ appear to be different. In the consensus half-site, the terminal nitrogen of Asn^{58} hydrogen bonds to P_E. In the nonconsensus half-site, the peptide -NH of Ala^{56} hydrogen bonds to $P_{E'}$.

The NH_2-terminal arm. Biochemical studies suggested that

repressor's NH_2-terminal arm contacts the major groove on the back of the operator site (*24*). When residues 1 to 3 were removed by tryptic digestion, the repressor was unable to protect guanines on the back of the operator from chemical methylation. Point mutations and deletions also suggested that this arm plays a critical role in recognition (*11, 25*). We found that the electron density for the first five residues, especially in the nonconsensus half-site, was not as clear as for other regions of the complex. We presume that the arms are partially disordered. In spite of these problems, density for the arm did appear as refinement proceeded, and we were therefore able to define the overall position of the arm.

As predicted, the arms wrap around the DNA, following the major groove to the center of the operator site (Fig. 2). In the consensus half-site, Lys^4 contacts the conserved guanine at base pair 6 (Fig. 3). The difference maps also show that this arm contacts base pair 8, and these contacts are made by Thr^2 in our model. The contacts made in the nonconsensus half-site are less well defined. However, the structure shows that the arms are not long enough to reach significantly beyond the center of the operator site. They cannot make contacts on the other half-site or interact with each other to form an antiparallel β sheet.

DNA conformation. As mentioned earlier, the operator site is a relatively straight segment of B-DNA. The average helical twist (34.4° per base pair) corresponds to an average of 10.5 bp per turn (Table 2). This is very close to the average expected for B-form DNA (*29*) and suggests that protein binding and crystallization have not distorted the DNA significantly. As has been observed in single-crystal studies of B-form DNA (*30*), the base pairs have a significant



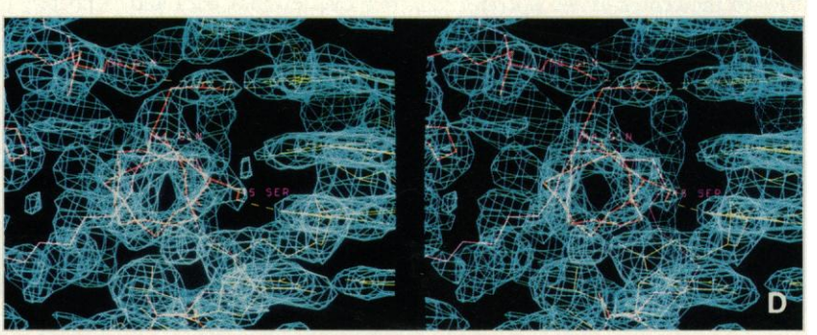


Fig. 3. (**A**) Stereo diagram showing the side chains that interact with base pairs in the major groove. The backbone of the protein is shown for residues 1 to 58, and the NH_2-terminal arm begins in the lower right corner of this figure. Critical side chains, and the backbone of helices 2 and 3, are emphasized with bold lines. (**B**) Sketch, in the same orientation as (A), showing hydrogen bonds in the consensus half-site. In the upper half of the figure, the major groove is readily visible for base pairs 1 to 3; in the lower half, the minor groove is closest to the viewer. (**C**) Sketches summarizing how side chains hydrogen bond to base pairs 2, 4, and 6 in the consensus half-site. The solid lines show the connection to the sugar phosphate backbone. Since Gln^{33} actually hydrogen bonds to the backbone, the sketch of base pair 2 shows the phosphate on the 5′ side of the adenine. (**D**) Calculated electron density from a $2|F_o| - |F_c|$ map in the vicinity of helix 3. The protein is shown in red, the DNA in light green, and the electron density in light blue. Residues 33 to 52 were omitted from the model used to calculate F_c and α_c. Orientation is similar to that in (A) and (B).

propeller twist (average = 11.6°). The individual helical twists range from 21.5 to 47.0°, and the individual propeller twists range from 4.6 to 21.2° (Table 2). Some of these variations may reflect sequence-dependent aspects of the lambda operator structure. However, there probably are some changes as the repressor binds because it enhances the reactivity of G 8′ toward chemical methylation (*31*). This base pair has the highest propeller twist in the operator site (21.2°) and the step between base pairs 7′ and 8′ has the highest helical twist (47.0°).

Although the ends of adjacent double helices have significant stacking interactions within the cocrystal, the junction between adjacent duplexes does not form a perfectly continuous double helix. The overhanging bases appear to form a Hoogsteen base pair, and there is a bend in the helix axis. Because this region is likely to be significantly affected by crystal packing forces, the terminal bases have been omitted from Table 2.

Agreement with chemical, biochemical, and genetic data. The cocrystal structure provides a solid basis for understanding and interpreting previous studies of the lambda repressor-operator complex. This agreement with the structural data confirms that it is reasonable to use mutagenesis and chemical protection data in modeling protein-DNA complexes. Strong effects on chemical reactivity or strong effects of surface mutations in the protein are good evidence for direct contacts. Specifically, we note that:

1) Ethylation interference experiments had implicated the same ten phosphates that repressor contacts (*31*), and these sites also are the regions where hydroxyl radical footprinting shows the strongest effects (*32*). The very weak hydroxyl radical footprint effects outside of the operator site do not appear to result from any direct contact. The small number of charged residues that interact with the phosphates is consistent with the estimate (from a thermodynamic analysis of the effect of salt on the dissociation constant) that repressor displaces only two to three cations when it binds to the operator site (*33*).

2) Repressor makes contacts in the major groove with each of the guanines that is protected from chemical methylation (*31*), and the one guanine with enhanced reactivity has an unusual conformation. No adenines, which react at the N3 position in the minor groove, were protected from methylation, and the structure confirms that repressor does not make any contacts in the minor groove.

3) The conserved AT base pair at position 2 and the conserved CG base pair at position 4 (which occur in each half of the six operator sites) are readily explained, since the repressor makes strong contacts with these base pairs. The λ cro protein also has a glutamine and a serine at the start of helix 3 and may make similar contacts with base pairs 2 and 4 in each half-site (*13*).

4) Almost all the repressor mutants are now readily explained. Perhaps the most interesting mutants are those that were not readily explained by the initial model of the repressor-operator complex (*3*) but now are explained by the cocrystal structure. For example, it now is clear why mutations changing Lys^4 or Tyr^{22} are so defective in operator DNA recognition. Most of the mutants that increase the affinity for operator DNA are also readily explained (*14*): When lysine is substituted for glutamic acid at position 34, modeling shows that the lysine can readily be rotated to contact the phosphate that precedes P_A. The methylene groups of the lysine may (like the methylene groups of Glu^{34}) interact with the thymine at base pair 1. When asparagine or serine is substituted for glycine at position 48, modeling shows that only minor adjustments are needed to accommodate the new side chain and allow hydrogen bonding to base pair 3. The one tight-binding mutant that is not readily explained is the glutamic acid to lysine change at position 83. When a lysine side chain is put at this position, the terminal nitrogen is more than 10 Å away from the DNA, and it is not clear why this mutant binds more tightly. It is possible that lysine somehow stabilizes the protein dimer.

5) As in our initial model of the repressor-operator complex, side chains that are implicated in positive control (*12*) are exposed on the surface of repressor that would be close to RNA polymerase when it is bound to the promoter P_{RM}. However, it will be necessary to look for secondary effects of the $Glu^{34} \rightarrow Lys$ mutant and the $Gly^{43} \rightarrow Arg$ mutant, since these might also introduce new contacts with the DNA backbone.

Comparison with other cocrystal structures. Lambda repressor is one of a set of proteins that use the helix-turn-helix motif for site-specific recognition (*7*). This structure has also been observed in the λ cro protein (*1*), the *Escherichia coli* CAP (catabolite gene activator protein) protein (*2*), the *E. coli* trp repressor (*4*), and 434 repressor-operator complex (*6*). A comparison of the lambda repressor and the 434 repressor helps us understand how the conserved helix-turn-helix unit is used in recognition.

Knowing the amino acid sequence of the 434 repressor made it possible for us to align it with the helix-turn-helix motif found in other bacterial and bacteriophage repressors (*7, 34*). Crystallographic studies later revealed that this region does form a helix-turn-helix structure in the 434 repressor (*6*), and it now is clear that these two repressors use the helix-turn-helix motif in roughly similar ways. In each case, the first two residues of the "recognition helix" (Gln^{44} and Ser^{45} for the lambda repressor; Gln^{28} and Gln^{29} for 434 repressor) make site-specific contacts with the edges of the base pairs that are exposed in the major groove. However, there is no simple pattern that aligns these residues with particular base pairs in the operator. In the 434 complex, residues 28 and 29 contact adjacent base pairs; in the lambda complex, there is an intervening base pair between the sites contacted by residues 44 and 45.

In several cases, corresponding residues also are used to make contacts with the phosphates. The conserved glutamine at the start of the helix-turn-helix motif (Gln^{33} in lambda repressor; Gln^{17} in 434 repressor) provides a striking example of a conserved contact. In each structure, the peptide -NH of this residue contacts a phosphate

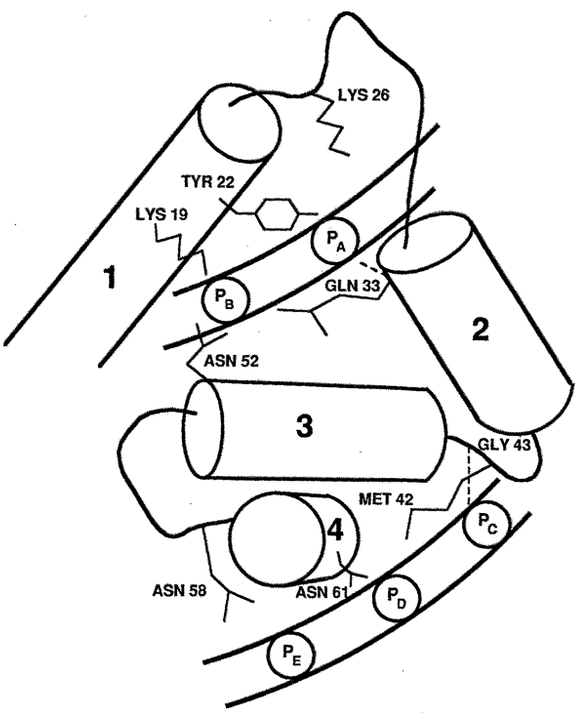


Fig. 4. Sketch showing side chains that interact with the sugar-phosphate backbone in the consensus half-site. View is at right angles to that in Figs. 2 and 3.

that flanks the operator site and the side chain hydrogen bonds to the glutamine at the start of the recognition helix. Contacts outside the helix-turn-helix unit seem to be less well conserved. Although residues in these flanking regions make important contacts with the DNA, these seem to have been less tightly constrained during evolution. They have adapted, with distinct sequences and distinct three-dimensional structures, to optimize recognition of their own operator sites.

The only other cocrystal structure of a site specific binding protein is the complex of Eco RI with its restriction site (*5*), and this structure is different from the repressor-operator complexes in several ways. (i) Eco RI does not use the helix-turn-helix motif for recognition, but has a set of α helices that are more nearly perpendicular to the major groove. (ii) Although the two proteins make about the same number of hydrogen bonds with the edges of the base pairs, the Eco RI restriction enzyme makes two hydrogen bonds with each of the six base pairs in the binding site. The λ repressor also makes about a dozen hydrogen bonds with the base pairs, but these are spread out over a 17-bp binding site. (iii) The Eco RI restriction enzyme distorts the DNA as it binds, whereas the lambda operator site is more nearly uniform B-DNA.

In large measure, these structural differences can be rationalized by considering the very different functions of a repressor and a restriction enzyme. Having a restriction enzyme make two hydrogen bonds per base pair may enhance specificity and help to protect against accidental cleavage at any other sites. By contrast, the lambda repressor must recognize a set of related sequences in the phage DNA, and there is no drastic penalty if repressor occasionally binds (nonspecifically) to some other sites in the phage or bacterial DNA. Distortion of the DNA may also play a functional role for Eco RI, possibly by enhancing the specificity of binding or preparing the DNA for cleavage. Evidently, significant distortion is not required as repressor binds to its operator. Other contacts are sufficient to provide the necessary level of sequence specificity, and significant distortions would merely lower the overall binding energy.

Implications for protein-DNA recognition. Previous studies of repressors (*1–4*) and of repressor-operator complexes (*6*) had given a preliminary picture of how these proteins bind to DNA, and our study confirms many of those basic conclusions. As expected, repressor binds as a dimer and uses the conserved helix-turn-helix unit to contact adjacent major grooves along one face of a B-form operator site. However, the lambda cocrystal structure gives us important new information about recognition.

An important observation from the cocrystal structure is that several side chains can "cooperate" to recognize a single base. This occurs in two different ways in the repressor-operator complex. At base pair 2, Gln^{33} plays a critical role even though it does not make any direct contact with the base pair (Fig. 3). Gln^{33} helps hold Gln^{44} in the right orientation for binding, and this explains why glutamine occurs so frequently as the first residue in the helix-turn-helix unit (*7*). The interaction of Gln^{33} with the phosphate oxygen will also help to precisely fix the orientation of Gln^{44}. This interaction between a side chain at the start of helix 2 and a side chain at the start of helix 3 complicates the interpretation of "helix swap" experiments that interchange the "recognition helices" of two repressors (*35*). If other repressors with glutamine at the start of the helix-turn-helix unit make similar contacts, these proteins would comprise a distinct subclass of the helix-turn-helix proteins, and "swap experiments" within this subset should have a greater chance of succeeding. [One very successful experiment spliced the recognition helix of the P22 repressor into the 434 repressor, and both of these helix-turn-helix units start with glutamine (*35*).]

At base pair 6, side chains also "cooperate" to enhance the specificity of recognition (Fig. 3), but in this case both side chains contact the base. Asn^{55} donates a hydrogen bond to the N7 of guanine, and Lys^{4} donates hydrogen bonds to the O6 of guanine and to the oxygen of the asparagine side chain. This cooperative interaction explains several otherwise puzzling observations, including the reduction in affinity observed when Lys^{4} is changed to glutamine (*11*) and the observation that a mutation at base pair 6 has a more drastic effect when the full amino-terminal arm is present (*25*). Like the glutamine-glutamine interaction discussed above, this interaction complicates efforts to use deletion experiments or swap experiments to determine how a particular residue or region of the protein affects recognition.

This cocrystal structure and the structure of the 434 complex (*6, 27*) both reveal an extensive network of interactions with the sugar-phosphate backbone. Presumably, these contacts help to precisely position the helix-turn-helix unit, and it is interesting that most of these contacts involve polar side chains or peptide -NH groups. Surprisingly, lysine and arginine are not used to make many of the phosphate contacts. Although they still could contribute to the overall binding energy, the loop containing Lys^{24}, Lys^{25}, and Lys^{26} only makes one tentative contact (at Lys^{26}) with the DNA. This apparent preference for small polar side chains could be significant for several reasons: The shorter side chains, which are less flexible, may help to position the repressor more precisely. It also is possible that hydrogen bonds with polar side chains may have a greater directional specificity than salt bridges with the charged groups of lysine or arginine. Hydrogen bonds (observed in several cases) that involve direct contacts between the protein backbone and the DNA backbone should help to position the helix-turn-helix unit very precisely.

It has been proposed that sequence-dependent variations in the structure of the DNA backbone may be important for site-specific recognition (*36*). However, in the lambda complex there are no dramatic kinks or distortions in the DNA, and it seems more likely that these backbone contacts enhance specificity by positioning the residues that contact the bases. These "nonspecific" contacts surround the "recognition helix," with residues at the beginning, middle, and end of the helix-turn-helix unit contacting the sugar-phosphate backbone (Fig. 4). The precise structural requirements for these contacts may help explain why this structural unit is so precisely conserved in a set of DNA-binding proteins (*1–4, 7*). Residues from the COOH-terminal end of helix 1 and the NH_2-terminal end of helix 4 also contact the DNA backbone. These contacts may explain a limited structural homology noted in the helix that precedes the helix-turn-helix unit (*37, 38*). However, differences in these flanking regions may allow different repressors to use the helix-turn-helix region in slightly different ways.

Our results emphasize the structural complexity of protein-DNA interactions. Although the overall orientation and some of the critical contacts were predicted correctly from a knowledge of the protein structure (*3*), many details of the interactions had not been anticipated. Comparing known cocrystal structures and predicted structures of other repressor-operator complexes suggests that there is little prospect of any simple recognition code. It does not appear there is any strict rule telling how a particular amino acid will be used or how a particular base will be recognized. As observed in this complex and in the Eco RI complex, several amino acids from different sections of the protein may cooperate to recognize a particular base pair, and it appears that recognition, like protein folding, involves cooperative interactions. The complexity of the interactions underscores the inherent limitations of using genetic and biochemical studies to make detailed structural predictions. It will be necessary to determine the structure of a series of protein-DNA complexes and carefully analyze the energy of these interactions to see whether any general principles of recognition emerge.

REFERENCES AND NOTES

1. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *Nature* **290**, 754 (1981); D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, B. W. Matthews, *ibid.* **298**, 718 (1982).
2. D. B. McKay and T. A. Steitz, *ibid.* **290**, 744 (1981); D. B. McKay, I. T. Weber, T. A. Steitz, *J. Biol. Chem.* **257**, 9518 (1982); I. T. Weber and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3973 (1984).
3. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982); M. Lewis, A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, C. O. Pabo, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 435 (1983).
4. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. Lawson, P. B. Sigler, *Nature* **317**, 782 (1985).
5. J. A. McClarin *et al.*, *Science* **234**, 1526 (1986).
6. J. E. Anderson, M. Ptashne, S. C. Harrison, *Nature* **316**, 596 (1985); J. E. Anderson, M. Ptashne, S. C. Harrison, *ibid.* **326**, 846 (1987).
7. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **53**, 293 (1984).
8. A. Klug and D. Rhodes, *Trends Biochem. Sci.* **12**, 464 (1987); J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
9. M. Ptashne, *A Genetic Switch* (Cell Press; Cambridge, MA, 1986).
10. C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1608 (1979); R. T. Sauer, C. O. Pabo, B. J. Meyer, M. Ptashne, K. C. Backman, *Nature* **279**, 396 (1979).
11. H. C. M. Nelson, M. H. Hecht, R. T. Sauer, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 441 (1983); M. H. Hecht, H. C. M. Nelson, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2676 (1983); H. C. M. Nelson and R. T. Sauer, *J. Mol. Biol.* **192**, 27 (1986).
12. A. Hochschild, N. Irwin, M. Ptashne, *Cell* **32**, 319 (1983).
13. A. Hochschild and M. Ptashne, *ibid.* **44**, 925 (1986); A. Hochschild, J. Douhan, M. Ptashne, *ibid.* **47**, 807 (1987).
14. H. C. M. Nelson and R. T. Sauer, *ibid.* **42**, 549 (1985).
15. S. R. Jordan, T. V. Whitcombe, J. M. Berg, C. O. Pabo, *Science* **230**, 1383 (1985).
16. J. L. Sussman, *Methods Enzymol.* **115**, 271 (1985).
17. W. Steigemann, thesis, Technische Universität, München (1974).
18. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
19. W. A. Hendrickson, *ibid.*, p. 252.
20. E. Westhof, P. Dumas, D. Moras, *J. Mol. Biol.* **184**, 119 (1985).
21. L. Beamer, S. Jordan, C. O. Pabo, unpublished data.
22. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
23. All data with $I > 2\sigma$ were used. The root-mean-square (rms) deviation for bond lengths was 0.015 Å, and the rms deviation for bond angles was 1.7 degrees.
24. C. O. Pabo, W. Krovatin, A. Jeffrey, R. T. Sauer, *Nature* **298**, 441 (1982).
25. J. L. Eliason, M. A. Weiss, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2339 (1985).
26. N. C. Seeman, J. M. Rosenberg, A. Rich, *ibid.* **73**, 804 (1976).
27. A. Aggarwal, D. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988).
28. W. G. J. Hol, P. T. van Duijnen, H. J. C. Berendsen, *Nature* **273**, 443 (1978).
29. J. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 200 (1979).
30. R. E. Dickerson and H. R. Drew, *J. Mol. Biol.* **149**, 761 (1981).
31. A. Johnson, thesis, Harvard University (1980).
32. T. D. Tullius and B. A. Dombroski, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469 (1986).
33. R. T. Sauer, thesis, Harvard University (1979).
34. R. T. Sauer, R. R. Yocum, R. F. Doolittle, M. Lewis, C. O. Pabo, *Nature* **298**, 447 (1982).
35. R. P. Wharton and M. Ptashne, *ibid.* **316**, 601 (1985).
36. R. E. Dickerson, *Sci. Am.* **249** (no. 6), 94 (1983).
37. T. A. Steitz, D. H. Ohlendorf, D. B. McKay, W. F. Anderson, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3097 (1982).
38. D. H. Ohlendorf, W. F. Anderson, M. Lewis, C. O. Pabo, B. W. Matthews, *J. Mol. Biol.* **169**, 757 (1983).
39. We thank Brien Byers, Van Vogel, and Clark Riley for help with the DNA synthesis; Timothy Whitcombe and Beishan Liu for help with purification and crystallization; Andy Howard for help with data collection and data reduction; Jeremy Berg for help with the preliminary analysis of crystal packing; Upul Obeysekare for help with programming and graphics; Lesa Beamer for permission to cite unpublished data; and Kathleen Kolish for help in preparation of this manuscript. Supported by NIH grant GM-31471, by the Howard Hughes Medical Institute, and by a contract from the Office of Naval Research. The coordinates have been deposited in Brookhaven Protein Data Bank.

11 July 1988; accepted 23 September 1988

Recognition of a DNA Operator by the Repressor of Phage 434: A View at High Resolution

ANEEL K. AGGARWAL, DAVID W. RODGERS, MARIE DROTTAR, MARK PTASHNE, STEPHEN C. HARRISON*

The repressors of temperate bacteriophages such as 434 and lambda control transcription by binding to a set of DNA operator sites. The different affinity of repressor for each of these sites ensures efficient regulation. High-resolution x-ray crystallography was used to study the DNA-binding domain of phage 434 repressor in complex with a synthetic DNA operator. The structure shows recognition of the operator by direct interactions with base pairs in the major groove, combined with the sequence-dependent ability of DNA to adopt the required conformation on binding repressor. In particular, a network of three-centered bifurcated hydrogen bonds among base pairs in the operator helps explain why 434 repressor prefers certain sites over others. These bonds, which stabilize the conformation of the bound DNA, can form only with certain sequences.

PROTEINS THAT REGULATE GENE EXPRESSION GENERALLY recognize specific DNA sequences through the binding properties of a distinct domain. Several classes of such domains have now been identified (*1, 2*). Members of the best studied class contain a "helix-turn-helix" element—a 20-residue motif of nearly invariant geometry but considerable sequence variation (*1, 3*). A first crystallographic analysis of a complex of one of these, the DNA binding domain of phage 434 repressor, with a synthetic DNA operator revealed an intricate combination of direct protein-DNA interactions and DNA conformational effects (*4*). That structure was determined from crystals diffracting to spacings of 3.2 Å in some directions but only to about 4.5 Å in others. Taken

The authors are in the Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138. A. K. Aggarwal and S. C. Harrison are in the Howard Hughes Medical Institute.

*To whom correspondence should be addressed.