Reprinted from Nature, Vol. 326, No. 6116, pp. 846-852, 30 April 1987

Structure of the repressor–operator complex of bacteriophage 434

J. E. Anderson, M. Ptashne & S. C. Harrison

Structure of the repressor–operator complex of bacteriophage 434

J. E. Anderson*, M. Ptashne & S. C. Harrison

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

The crystal structure of a specific complex between the DNA-binding domain of phage 434 repressor and a synthetic 434 operator DNA shows interactions that determine sequence-dependent affinity. The repressor recognizes its operators by its complementarity to a particular DNA conformation as well as by direct interaction with base pairs in the major groove.

WE describe here the structure of a specific complex between the DNA-binding domain of a bacteriophage repressor protein and a synthetic operator DNA. The description is based on an X-ray crystallographic analysis of crystals that diffract to 3.2 Å resolution in some directions and to about 4.5 Å in others. The repressor is encoded by coliphage 434, a close relative of phage λ. It binds to a set of six similar but non-identical 14-base-pair (bp) operators in the phage genome. Its differential affinity for these sites, coupled with the distinct differential affinity of the homologous cro protein, creates a regulatory switch determining the choice between lysogeny and lytic growth[1]. The structure of the repressor–operator complex, taken together with experiments reported in an accompanying paper and elsewhere[2,3], reveals the basis of specific binding and its modulation by the exact DNA sequence of the operator.

In an account of this structure at 7 Å resolution[4], we described the general features of the complex. The repressor DNA-binding domain (R1-69), contains the first 69 residues of the complete polypeptide. Its conformation is predominantly α-helical, similar to the first four helices of the λ repressor[5]. The second and third alpha helices (α2 and α3) form a helix–turn–helix motif, found in crystal structures of certain other DNA binding proteins[6–8] and believed on the basis of amino-acid sequence similarity to occur in many more[9]. The conformation of the 14-bp synthetic operator (14-mer) is B-DNA-like, and individual 14-mers are stacked end-to-end to form pseudocontinuous double helices running through the crystals. Two R1-69 subunits are bound to each 14-mer, and α3 of each monomer rests in the major groove of a half-site.

The higher resolution structure presented here shows the interaction in molecular detail. Hydrogen bonds between amide nitrogens of the peptide backbone and phosphate groups on DNA strongly constrain the fit of DNA to protein, and three glutamine side chains on α3 are in position to form a pattern of van der Waals and hydrogen-bond contacts with the five outer base pairs. Thus, the repressor recognizes in DNA both a particular conformation and an array of base-pair contacts.

Structure determination

Crystals of R1-69 and the 14mer were prepared as described previously[10]. Data-collection procedures and statistics are presented in Table 1. Initial single isomorphous replacement (SIR) phases were obtained using crystals with DNA containing 5-bromodeoxyuracil instead of thymine at position 7. Phases between 15 Å and 7 Å were refined using real space noncrystallographic symmetry averaging[11–13] about a local 3_1 axis parallel to [1$\bar{1}$1] and intersecting [110] 33.5 Å from the origin[4]. The small magnitude of the heavy-atom differences and weak intensities beyond 7 Å made conventional refinement of heavy-atom positions impractical. We therefore used a two-step strategy to extend phases, first from 7 Å to 5 Å and then from 5 Å to 3.2 Å. In the first step, heavy-atom positions were refined at 7 Å resolution by averaging a 15–7 Å heavy-atom difference map about a series of local 3_1 axes, parallel to [1$\bar{1}$1] and intersecting [110] in the neighbourhood of 33.5 Å from the origin. The position of the 3_1 axis that gave the strongest average heavy-atom peaks was selected, and new heavy-atom positions were determined by inspection of the corresponding map. SIR phases between 7 and 5 Å resolution were computed using these positions, combined with the previously refined 15 to 7 Å phases, and used to generate a 5 Å map. Eight cycles of noncrystallographic symmetry averaging[13] about the new local 3_1 (34.0 Å from the origin along [110]) gave an averaging R-factor, R_{av} = 0.28. In the second step, refined phases to 5 Å resolution were used in an analogous way to adjust the origin of the threefold screw axis and to improve the heavy-atom coordinates. The noncrystallographic symmetry phase refinement was unstable if all phases from 15 to 3.2 Å were allowed to change, but fixing phases from 15 to 5 Å at their best values from the 5 Å refinement allowed the computation to converge with R_{av} = 0.30 for all data from 15 to 3.2 Å (Table 2, upper part). This way of refining heavy-atom coordinates is similar to the procedure introduced by Hogle for work on poliovirus[14].

The map thus computed showed clear double-helical DNA density, each deoxyribose-phosphate backbone having a staircase like shape, with sugars corresponding to the risers. Base pairs were poorly defined by density, but an unambiguous model could be built using backbone density and hydrogen-bonding restraints as a guide. Four α-helices of the protein appeared as twisted rods. A number of side chains were well-defined; others appeared as truncated bulges. For model-building, we used the program FRODO[15] with separate files for DNA and protein coordinates. The 14mer was broken into single stranded mono-, di- and trinucleotide segments, the phosphates and sugars were fitted to the DNA backbone density, and the torsion angles were adjusted to restore approximately correct base-pair geometry. The Hendrickson–Konnert restrained least squares refinement programs[16], modified by G. Quigley (MIT) to accept nucleic acid coordinates, were used with base-pair hydrogen bonding restraints to restore idealized geometry. Further manual adjustment followed, with periodic regularization. Residues 1–58 of the protein were built into the map and adjusted with the regularization routines in FRODO. Finally, protein and DNA coordinates were regularized with the constrained–restrained least squares refinement program CORELS[17], with base-pair hydrogen bonding restraints imposed on the DNA.

The initial model, constructed as just described, was used to obtain starting phases for a second round of noncrystallographic symmetry phase refinement. The order in the crystals is anisotropic, with intensities measurable to spacings of 3.2 Å in the directions of noncrystallographic symmetry axes formed by stacked protein–DNA complexes, but only to about 4.5 Å in other directions[10]. The map can therefore be regarded as density at 3.2 Å resolution, selectively 'smeared' in directions perpendicular to the non-crystallographic symmetry axis. When build-

* Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, Long Island, New York 11724, USA.

Anderson et al (1987) Nature 326, 846-852.

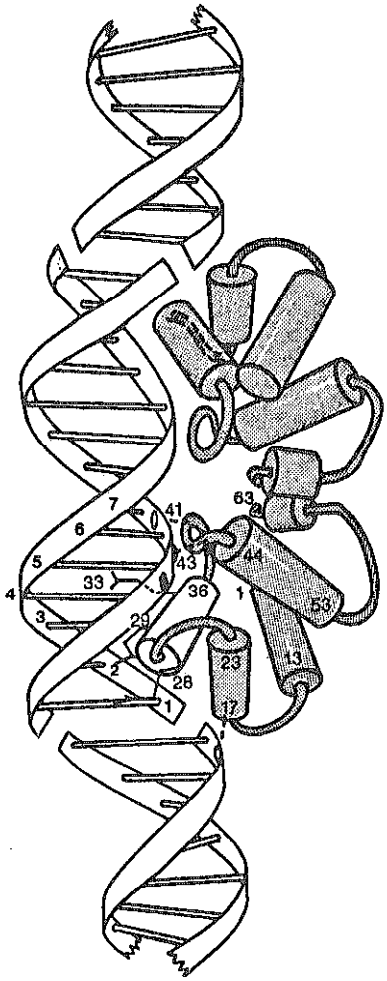
Please note the following corrections:

p. 849 (5 lines from bottom of first column): arg 10 (not 5)

Fig. 5b: O4 (not O6) on T of base pair 4

Table 3: Residue 13 (not 15) of λ repressor corresponds in three-dimensional position to residue 1 of 434 repressor.

Refined high-resolution structures of R1-69 (2Å) and 434 cro (2.5Å), now complete, show the C terminal region more accurately. Residues 57-61 actually form a short fifth helix, and residues 64-69 are disordered, even in the free proteins (Mondragon, Subbiah, Almo, Wolberger and Harrison, unpublished). An updated version of the diagram in Fig. 1a, reflecting this improved model, is shown below. The fit to density in the map of the R1-69/14mer complex is very good.



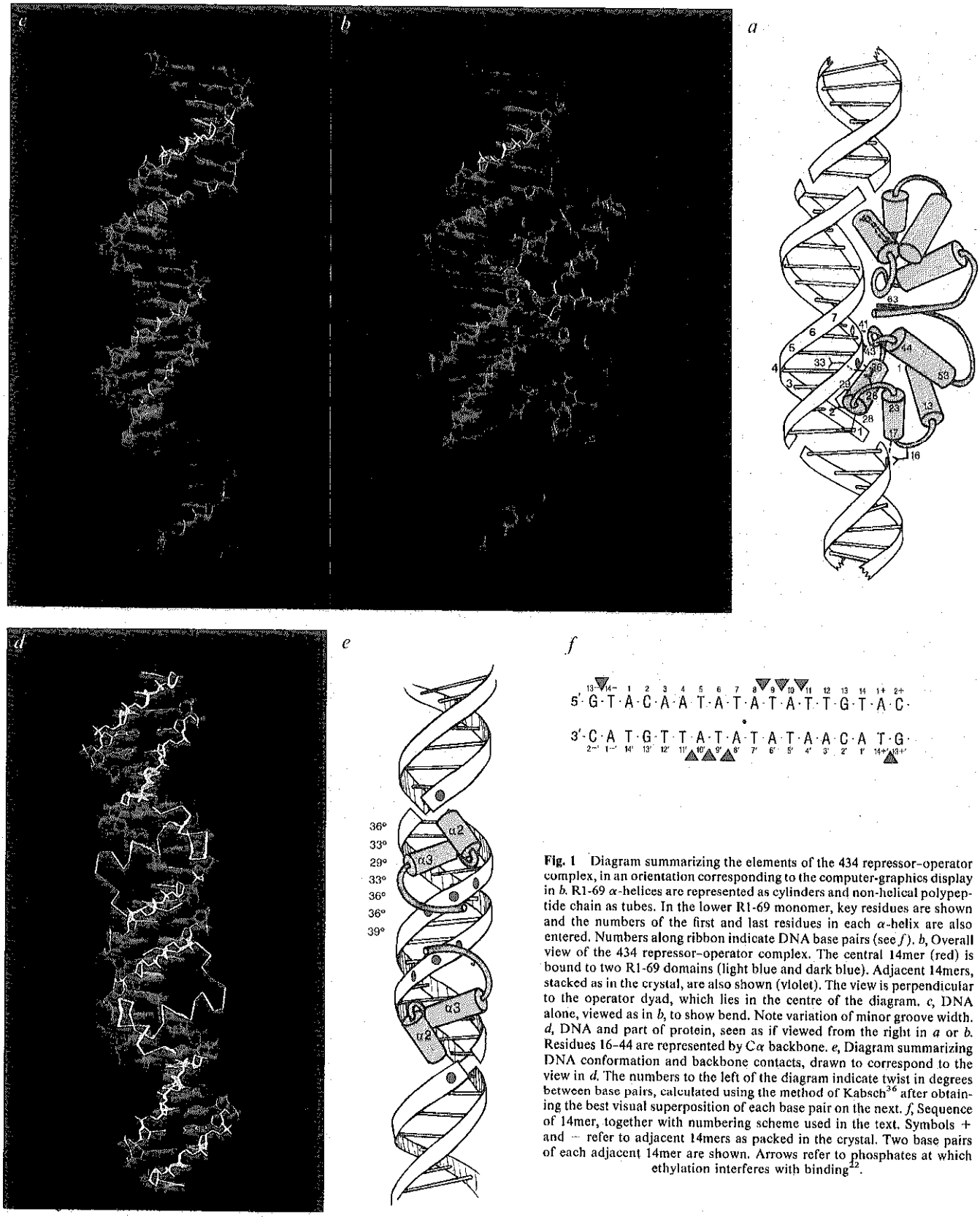


Fig. 1 Diagram summarizing the elements of the 434 repressor–operator complex, in an orientation corresponding to the computer-graphics display in *b*. R1-69 α-helices are represented as cylinders and non-helical polypeptide chain as tubes. In the lower R1-69 monomer, key residues are shown and the numbers of the first and last residues in each α-helix are also entered. Numbers along ribbon indicate DNA base pairs (see *f*). *b*, Overall view of the 434 repressor–operator complex. The central 14mer (red) is bound to two R1-69 domains (light blue and dark blue). Adjacent 14mers, stacked as in the crystal, are also shown (violet). The view is perpendicular to the operator dyad, which lies in the centre of the diagram. *c*, DNA alone, viewed as in *b*, to show bend. Note variation of minor groove width. *d*, DNA and part of protein, seen as if viewed from the right in *a* or *b*. Residues 16–44 are represented by Cα backbone. *e*, Diagram summarizing DNA conformation and backbone contacts, drawn to correspond to the view in *d*. The numbers to the left of the diagram indicate twist in degrees between base pairs, calculated using the method of Kabsch[36] after obtaining the best visual superposition of each base pair on the next. *f*, Sequence of 14mer, together with numbering scheme used in the text. Symbols + and − refer to adjacent 14mers as packed in the crystal. Two base pairs of each adjacent 14mer are shown. Arrows refer to phosphates at which ethylation interferes with binding[12].

ing a model, however, stereochemical constraints couple positions along the local symmetry axis with those normal to it. Therefore, the effective resolution is intermediate between the two limits. To 5 Å, the intensity fall-off is relatively isotropic and computed structure factors at this resolution were used in a translation search program (A. Aggarwal, personal communication) to carry out a final adjustment of the position of the 3_1 axis. The anisotropy and non-crystallographic symmetry were incorporated into FFT (fast Fourier transform) structure factor calculations[18] by using the Bricogne density reconstitution procedure[13] to generate a full crystallographic asymmetric unit from a half-complex 'map' that is computed in a coordinate frame with the noncrystallographic symmetry axis along *Z*. This choice of frame simplified introduction of a uniaxial, anisotropic 'tem-

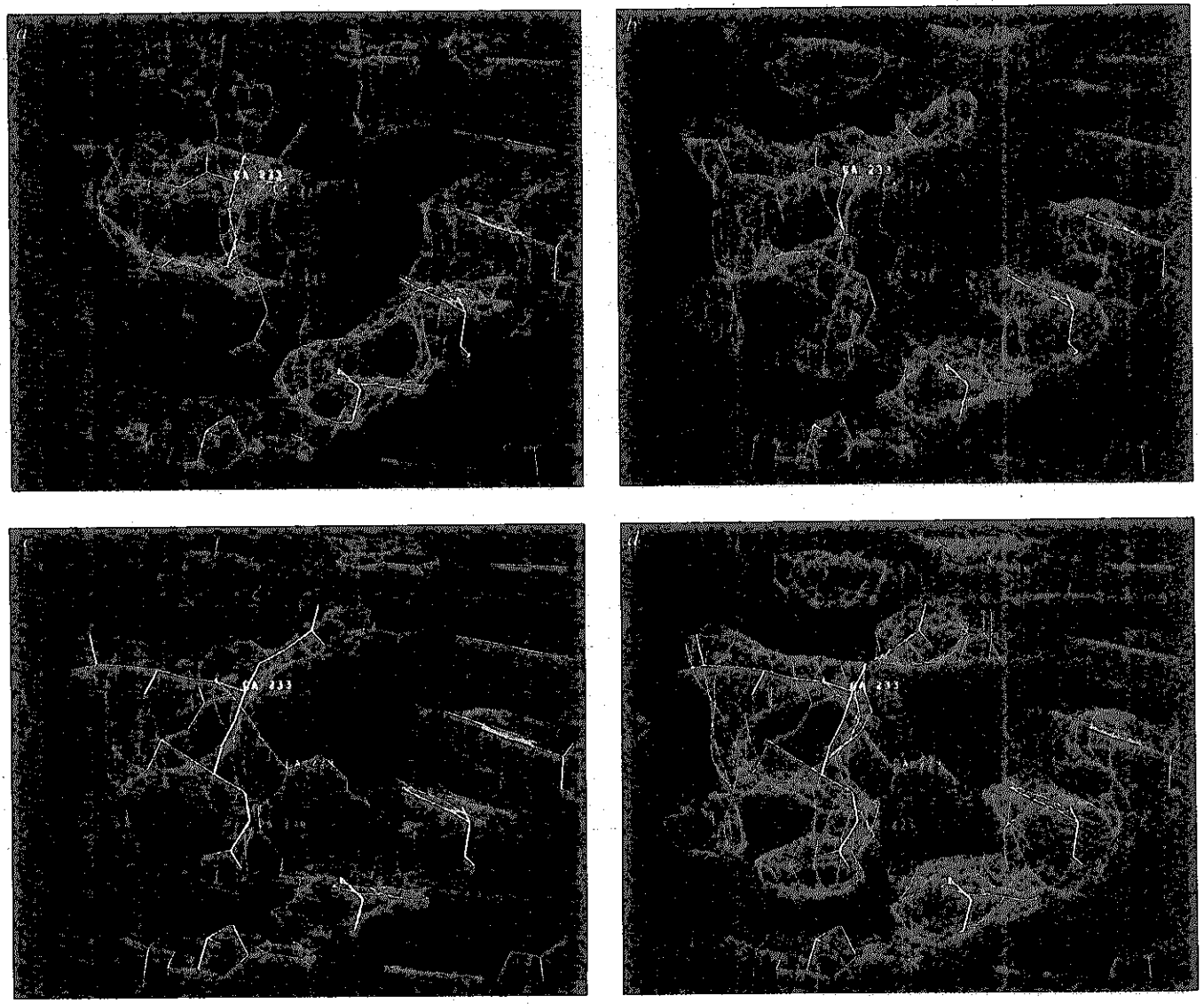


Fig. 2 Part of the electron density maps at different stages of the structure determination, showing the N-terminal half of α3 and portions of base pairs 1–4. Models are superimposed on density. The helix α3 is viewed from the C-terminal end, with residues 28–34 (QQSIEQL) shown and α-carbons of residues 28, 29 and 33 labelled (228, 229 and 233, respectively). *a*, Map after refinement of SIR phases, with initial model superimposed. *b*, Map after refinement of phases determined from initial model, with initial model superimposed. Note clear indications for repositioning of side chains, especially Gln 28 and Gln 29 (background) and Gln 32 (foreground). *c*, Same map as in *b*, but with rebuilt model. *d* Final $(2F_o - F_c)$ map, with rebuilt model.

perature factor'. Various ratios of Bx = By to Bz were examined, by comparing calculated and observed amplitudes. The optimum occurred at 1.75, as judged both by R factor minimum and by relative lack of noise in corresponding $2F_o - F_c$ and $F_o - F_c$ maps.

Phases from the initial model were used to initiate eight further cycles of noncrystallographic symmetry refinement of phases between 15 and 3.2 Å. The computation converged stably to $R_{av} = 0.27$, with an overall phase change of 43° (Table 2, lower part). Significant adjustments were indicated by this map, and we extended the model from residue 58 to residue 63 (Fig. 1). Two further rounds of smaller adjustments were based on $2F_o - F_c$ and $F_o - F_c$ maps. Maps and coordinates at various stages are shown in Fig. 2. It is evident that the second round of non-crystallographic-symmetry refinement substantially improved phase determination. Base-pair and side-chain densities became far clearer and peptide-backbone density more continuous. Poor placement of some side chains of the initial model did not prevent appearance of density in correct positions in the final map, showing that errors in that model did not unduly bias the outcome of the phase refinement (Fig. 2*b* and *c*). The use of an intermediate model to initiate a second round of phase refinement was used in determining the turnip crinkle virus structure, where similar corrections to an actual structure were generated[19]. The power of threefold non-crystallographic redundancy in phase refinement has been amply demonstrated in previous work—for example, in the determination of the influenza virus haemagglutinin structure, with data of comparable accuracy to ours[20]. Reciprocal-space refinement of the model is in progress. The R-factor of the model presented here is 0.44. (H. Holley, J.E.A. and S.C.H. unpublished data).

As in any structure determination at this resolution, interactions such as hydrogen bonds must be inferred from the position and orientation of the participating groups. We believe our present structure to be sufficiently well determined that the protein–DNA interactions can be described correctly and that likely hydrogen bonds can be assigned. More precise conformational details, accurate hydrogen-bond geometry, contact distances and features such as the positions of water molecules—all of which may be significant for complete understanding of specificity—will only be visible at the higher resolution afforded by a different R1-69–operator crystal currently being studied (A. Aggarwal and S.C.H. unpublished data).

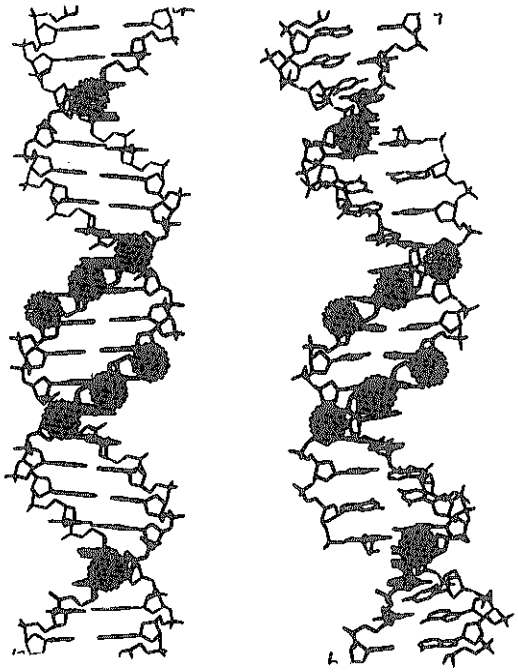


Fig. 3 Conformation of DNA in the crystalline complex (right) compared with idealized, 10.5-bp-per-turn B-DNA (left). Contacted phosphates are highlighted. It is clear from this figure, and from Fig. 1, that model building with idealized B-DNA could not correctly predict the protein–DNA contacts.

DNA conformation

The DNA of the 14mer forms a B-type helix throughout its length. The 14mers are stacked accurately on adjacent DNA segments to form a pseudocontinuous helix. The local twist varies from about 39° to 29° per base pair—that is, from 9.2 to 12.3 base pairs per turn (Fig. 1*e*). A similar range of variation in twist has been found in crystalline B-DNA[21]. With respect to the average of 10.5 base pairs per turn, the 14mer is overwound at its centre and underwound at its ends. The helix axis bends somewhat, causing the DNA to curl slightly around helix 3 of each monomer. The bend is sharpest between base pairs 4 and 5 (and 10 and 11), with relatively more gentle bending between base pairs 1 and 4, between base pairs 5 and 10, and between base pairs 11 and 14. Overwinding and bending in the centre of the DNA fragment narrows the minor groove. Underwinding and bending widens it at the ends. This variation is evident in Figs 1 and 3.

Protein conformation

R1-69 is a cluster of four α-helices, with a C-terminal extension. Two of the helices, $\alpha 2$ and $\alpha 3$, form a helix-turn-helix motif. The residues found in each of the four helices are shown in Table 3. For comparison, the corresponding residues in the N-terminal domain of λ repressor[5], determined by comparison of the two models, are also listed. As observed in our earlier paper[4], if the 434 and λ $\alpha 2$-$\alpha 3$ structures are superposed, the axes of $\alpha 1$ and $\alpha 4$ in the two repressors coincide to within about 2–3 Å. The α-carbon positions suggest that 434 repressor has 'deletion' with respect to λ repressor of two residues between $\alpha 1$ and $\alpha 2$ and of one residue at the beginning of $\alpha 4$. The two models diverge at residue 58 of 434 (corresponding to 75 of λ), precisely the point at which the two amino-acid sequences cease to be similar. Various hydrophobic side-chains form the interior of the four-helix cluster, which is also stabilized by several polar linkages—notably, Gln 17–Glu 32 and Arg 5–Glu 35. The last six residues of R1-69 are poorly defined in our map. They appear to extend toward DNA backbone near P11.

Interactions occur between the two R1-69 subunits bound to one 14mer. The two monomers are in contact across the dyad near residues 57–58. The aliphatic chain of Arg 41 and a residue near the C-terminus in one subunit lie against the aromatic ring of Phe 44 in the other subunit. These contacts may be important in determining the relative orientations of the DNA-binding surfaces of the two monomers, thereby defining the required spatial orientation of one operator half-site with respect to the other.

Table 1 Statistics for data between 45 and 3.2 Å resolution

Native				
d_{min} (Å)	$\langle F \rangle$	$R_{sym}(N_{mult})$	N (% $F>2\sigma$)	N_{poss}(% $F>2\sigma$)
9.90	7,568	0.080 (148)	368 (98.4)	622 (58.2)
7.08	6,559	0.096 (585)	929 (98.5)	622 (58.2)
5.81	3,476	0.107 (721)	1,099 (97.0)	1,260 (84.6)
5.04	2,916	0.138 (815)	1,283 (95.9)	1,466 (83.9)
4.51	2,955	0.157 (920)	1,399 (94.1)	1,651 (79.8)
4.12	3,022	0.170 (951)	1,535 (94.2)	1,817 (79.6)
3.82	2,695	0.212 (982)	1,675 (92.0)	1,961 (78.6)
3.58	2,491	0.242 (941)	1,705 (88.7)	2,090 (72.3)
3.37	2,161	0.284 (921)	1,756 (82.8)	2,237 (65.0)
3.20	2,535	0.212 (834)	1,708 (79.7)	2,328 (58.5)
Overall	3,129	0.161 (7,818)	13,457 (90.7)	16,444 (74.2)
Derivative				
d_{min} (Å)	$\langle F \rangle$	R_{sym} (N_{mult})	N (% $F>2\sigma$)	N_{poss}(% $F>2\sigma$)
9.90	7,194	0.041 (59)	229 (97.4)	622 (35.9)
7.08	6,225	0.047 (240)	664 (96.8)	1,013 (63.5)
5.81	3,341	0.093 (387)	866 (88.6)	1,260 (60.9)
5.04	2,879	0.135 (442)	970 (84.7)	1,466 (56.1)
4.51	2,902	0.179 (471)	1,079 (85.1)	1,651 (55.6)
4.12	2,905	0.212 (424)	1,132 (83.7)	1,817 (52.2)
3.82	2,790	0.282 (366)	1,244 (79.7)	1,961 (50.6)
3.58	2,630	0.313 (274)	1,200 (75.5)	2,090 (43.3)
3.37	2,430	0.274 (244)	1,141 (70.9)	2,237 (36.2)
3.20	3,405	0.194 (261)	1,089 (75.0)	2,328 (35.1)
Overall	3,224	0.169 (3,168)	9,614 (81.6)	16,444 (47.7)

Mean isomorphous difference $= \sum |F_{nat} - F_{der}| / \sum |F_{nat}| = 0.22$, calculated after exclusion of data (between 45 and 3.2 Å) for which $|F_{nat} - F_{der}| > 5\sigma(\Delta F)$.

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

N_{mult} = number of reflections measured more than once; only these contribute to R_{sym}.

N = number of reflections measured (after rejections) including those measured only once.

N_{poss} = maximum number of reflections possible in corresponding resolution range.

The percentage of observed and possible data with $F > 2\sigma(F)$ is indicated. Data from native crystals were collected to 3.2 Å using CEA-25 X-ray film and oscillation photography with Elliot GX-6 and GX-13 X-ray generators. The films were scanned with an Optronics P1000 film scanner on a 50 μm raster. Integrated intensities were obtained with the film-scanning program SCANFILM, a derivative of SCAN12 (ref. 34). Data to 7 Å were also collected on a prototype Xentronics area detector with the Harvard software package[35]. These measurements provided intensities for native reflections that were too bright to be measured on film. Derivative data to 3.2 Å were collected on the area detector from isomorphous crystals prepared with DNA containing 5-bromodeoxyuracil instead of thymine at position 7 of the 14mer[4]. The native data from detector and film were processed and merged, then scaled to the processed derivative data. All programs for data processing and scaling were from P. Evans (MRC Laboratory of Molecular Biology).

Protein–DNA contacts

The polypeptide chain is so folded that in the complex the N-termini of α-helices 2, 3 and 4 all point toward the DNA backbone, $\alpha 3$ lies in the major groove, and the loop joining $\alpha 3$ and $\alpha 4$ runs along the DNA backbone (Fig. 1*d*). Contacts to DNA backbone from one R1-69 subunit occur on both sides of the major groove that is occupied by $\alpha 3$ and possibly across the minor groove that contains the operator dyad. Major-groove interactions with base pairs 1–5 and 10–14 are made by three glutamine side chains projecting from $\alpha 3$. These relationships are summarized in Figs 1*d* and *e*, and described in detail below.

Backbone contacts. Four phosphate groups per half-site, corresponding precisely to the phosphates implicated by ethylation interference experiments[22], appear to interact with protein. Considering half of the complex, the contact phosphate groups lie

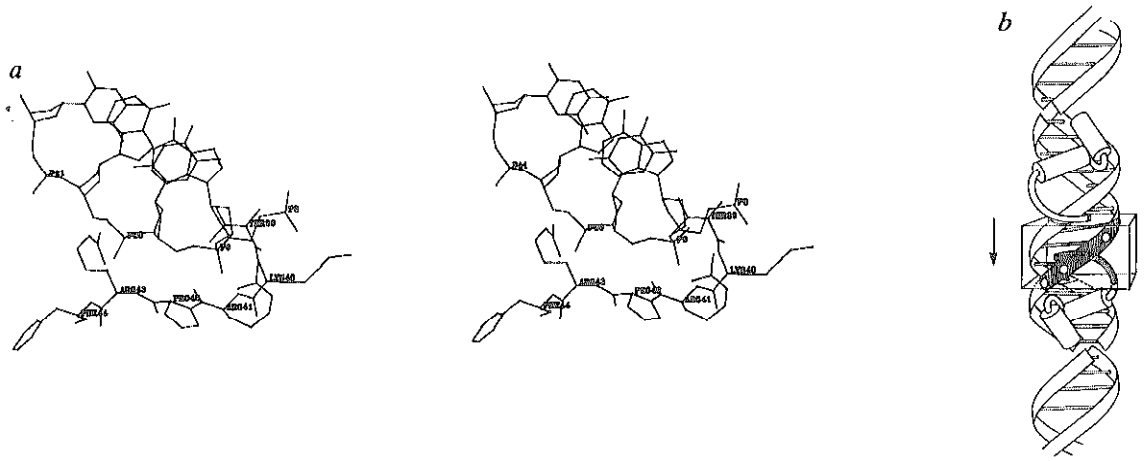


Fig. 4 Stereo view of contacts between protein and DNA phosphates P9′ and P10′. *b*, Schematic diagram as a key to *a*. The view is of the region indicated in the (shaded structures), looking approximately along the DNA axis (arrow). Compare key with Fig. 1*e*. Residues 39–44 of the protein are shown, together with nucleotides 8′–11′. (Primes are omitted in the labels.)

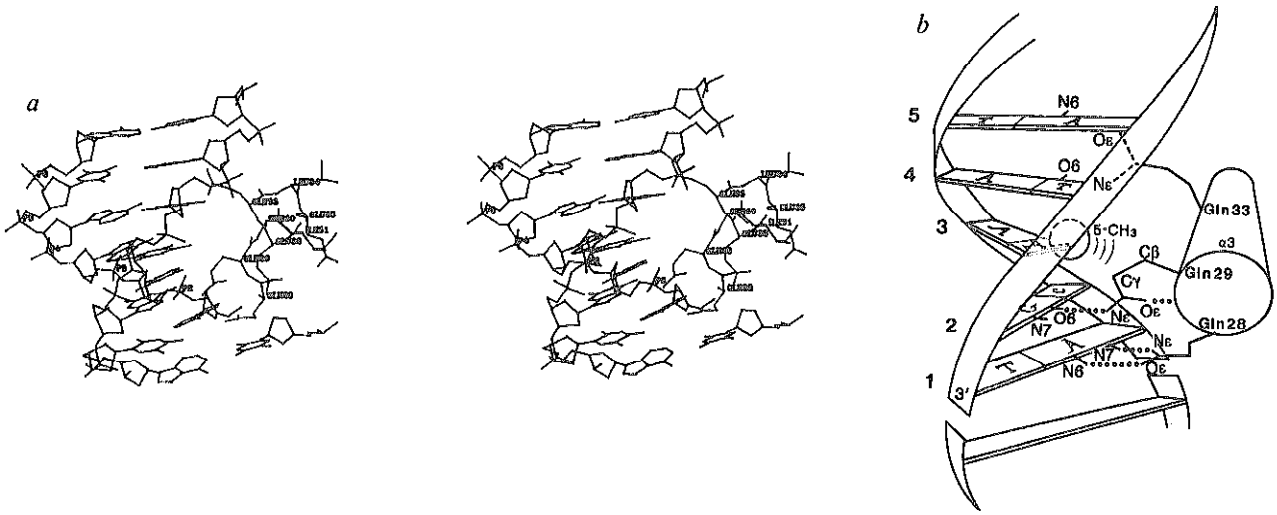


Fig. 5 *a*, Stereo view of helix α3 and base pairs 1–5, viewed from the N-terminus of α3. One base pair of the adjacent 14-mer is also present. Included are residues 28–35 of R1-69. *b*, Schematic diagram as a key to *a*. Note the similarity to view of α3 and surrounding structure in the lower monomer of Fig. 1*c*. Gln 28, Gln 29 and Gln 33 are shown, with hydrogen bonds suggested by the current model drawn in dotted lines. Functional groups on bases and on Gln 33 that might participate in additional hydrogen bonds are also shown. The proposed non-polar contact between the 5-CH_3 group of T12′ (base pair 3) and Cβ and Cγ of Gln is indicated.

5′ to bases 14-, 9′, 10′ and 11′ (see Fig. 1*f* for numbering scheme). The P14- contact occurs between different complexes in the crystal. Because the stacking of 14mers mimics continuous DNA, we believe that the interaction we observe at P14- closely reflects the one actually made in continuous DNA.

The three tightest contacts (as judged both from the structure and by the strength of ethylation interference) all appear to involve peptide backbone NH groups. At P14-, the NH of Gln 17 approaches the phosphate oxygens closely enough to form a hydrogen bond. Residue 17 is at the N-terminus of α2, and its peptide NH group is expected to make a particularly strong hydrogen bond to a negatively charged ion, as a result of oriented peptide dipoles in the α-helix[23]. Analogous interactions occur in the *Salmonella typhimurium* sulphate binding protein, where peptide NH groups at the N-termini of α helices are ligands of bound SO_4^{2-} ions[24]. The side chain of Asn 16 also appears to be involved in the contact to P14-. At P9′, there is close approach of the main chain NH groups of Lys 40 and Arg 41; the side-chain conformation of Arg 41 permits a salt link with the same phosphate (Fig. 4). At P10′, there is close approach of the peptide NH groups of both Arg 43 and Phe 44 (Fig. 4). Phe 44 is the amino-terminal residue of α4. The use of peptide NH groups to contact phosphates may be a common feature in proteins that bind DNA and nucleotides. Dreusicke and Schulz[25] have described a 'giant anion hole' at the site in adenylate kinase where a nucleotide phosphoryl group is believed to be located. The polypeptide chain loops in such a way that five peptide NH groups coordinate a bound sulphate in their crystal structure.

The side chain of Arg 43 projects into the minor groove near base pair 7. The groove is somewhat narrower than average at this position, and the guanidinium group could probably have significant coulombic interaction with phosphates 9 and 10′ on either side of the groove (Fig. 4). The possibility of direct interaction with base pair 7 is discussed in the next section and in the accompanying paper.

The ethylation of P11 interferes more weakly with binding than ethylation of P14-, P9′ or P10′ (ref. 22). Thr 26 appears to contact P11′, and the C-terminal residues of the other monomer may also lie in this vicinity. Because the C-terminal residues are not well defined in our map, we cannot at present determine whether any actually engage in backbone contacts. If they do, the interaction may help establish the favoured relative orientation of the two operator half-sites, because an R1-69 domain bound in one half-site would make contact by its C terminus with DNA backbone in the other half-site.

Base-pair contacts. Three glutamines (residues 28, 29 and 33) project from α3 into the major groove (Fig. 5). Gln 28

approaches A1 (base pair 1) in such a way that hydrogen bonds can readily form between N7 of the base and Nε of the glutamine and between N6 and Oε. 'Bidentate' H-bonding of this type was suggested as a mode of adenine-specific interaction by Seeman *et al.*[26]. Gln 29 lies along α3. Its Cβ and Cγ are in van der Waals contact with the 5-methyl group of T12′ (base pair 3), and its Nε can hydrogen bond to O6 (and perhaps also N7) of G13′ (base pair 2). The Oε of Gln 29 may hydrogen bond to its main chain NH, an interaction occasionally seen with glutamine at amino termini of α-helices[27]. This hydrogen bond would fix the orientation of the glutamine and point the Nε donor group directly at G13′. Gln 33 projects toward T11′ and A10′ (base pairs 4 and 5). There could be a hydrogen bond between Nε of Gln 33 and O4 of T11′ (base pair 4), and perhaps between Oε and N6 of A10′ (base pair 5); a refined model will be required for confident assignment.

The present model shows no direct interactions with base-pairs 6 and 7. Although residues 64–69 are not well defined in the map, they are clearly excluded from the major groove on the near side of the operator by α3 and from the minor groove by residues 40–44. Moreover, model building indicates that there is no reasonable possibility of base-pair contact from these residues in the major groove on the far side of the DNA.

Contributions to specificity

Three different features of the DNA structure appear to be significant for specificity: (1) the groups accessible in the major groove on base pairs 1–5, (2) the conformational details of the half-sites, and (3) the overwinding of the central base pair steps with a narrowing of the minor groove. These features are relevant to different classes of interactions made by the protein. (1) Side chains from α3 contact base pairs in the major groove. (2) Peptide amino groups, as well as side chains from various residues, contact backbone phosphates in ways that constrain the conformation of the operator half-sites. (3) Dimer interactions between repressor monomers favour a particular spatial relationship between half-sites.

Base-pair contacts. The interactions of Gln 28 with base pair 1 and of Gln 29 with base-pairs 2 and 3 are consistent with the conservation of base pairs 1–3 in all 12 operator half-sites, because changes in any of these base pairs would destroy one or more of the hydrogen bonds or disturb van der Waals contacts with the thymine methyl group. (For a list of all naturally-occurring operator sequences, see accompanying paper, ref. 2.) Wharton[3] used a random mutagenesis scheme to generate all possible changes in amino acids 28 and 29. No viable mutants were detected, and indeed no likely substitutions in the model can be imagined. Wharton and Ptashne[28] have also shown that a mutant operator with T instead of A at position 1 is recognized specifically by a mutant repressor with alanine at position 28. These coordinated changes can be built into the model, generating a van der Waals contact between the $-CH_3$ group of T1 and the $-CH_3$ side chain of Ala 28.

The side chain of Gln 33 is in position to contact base pairs 4 and 5. The model shows that this residue could hydrogen bond to either base 11′ or 10′, or to both (Fig. 5). Wharton[3] has studied the effects of substituting Ala for Gln 33 on the binding of repressor to various operators. His results, taken with others (R. Wharton and M.P. unpublished data), are consistent with the idea that residue 33 indeed contacts positions 4 and 5.

Backbone contacts. Within a half-site, contacts to the DNA backbone position a subunit of the protein against the operator. The relationship is likely to be quite exact, because main-chain peptide amino groups participate in the contacts and the main chain of a folded protein is generally less flexible than the side chains. The backbone contacts, especially the one at P14-, require a gently bent outer operator[22]. This curvature is essential for at least one of the major groove contacts: without it, the interaction between Gln 28 and A1 would not have the correct geometry. It is not yet clear whether the bend is present in unbound operator DNA, or whether protein binding induces it.

Central base pairs. The configuration of the operator at its central base pairs is consistent with sequence preferences at positions 6–9. A more detailed discussion appears in the accompanying paper, but we summarize the arguments here. The twist is about 36° between base pairs 6 and 7 and base pairs 8 and 9 and 39° between base pairs 7 and 8 (Fig. 1*e*). The minor groove is consequently 2.5 Å narrower than in idealized B-DNA. It is believed that GC or CG base pairs are not well accommodated in a narrow minor groove[29], and repressor indeed discriminates against GC or CG at positions 6–9 (ref. 2). The sequence preferences at central base pairs persist when alanine is substituted for Arg 43, which projects into the minor groove. The position of the guanidinium group in the present model does not indicate hydrogen bonding to bases, and the binding affinities of the Ala 43 mutant show that any minor-groove contacts that are made do not contribute significantly to sequence discrimination at positions 6–9. We therefore believe that it is the DNA conformation, rather than contacts to the bases themselves, that influences binding.

How does the structure in the centre of the operator influence affinity? One explanation is that a particular conformation is required for correct relative orientation of the two half-sites. Though monomeric at most concentrations in solution, the two R1-69 domains in the complex with DNA have significant interactions, including complementary packing of parts of each subunit that are important for DNA backbone contacts (especially residues 41–44: see Fig. 4). Thus, the relative orientation of

Table 2 Phase refinement by noncrystallographic symmetry averaging

Refinement of SIR phases*

	Phases refined 15→3.2 Å			Phases refined 5→3.2 Å		
Cycle	R_{av}	r	Mean phase change	R_{av}	r	Mean phase change
1	0.58	0.537	53.3	0.55	0.346	75.3
2	0.45	0.734	40.1	0.41	0.600	34.7
3	0.39	0.785	21.9	0.35	0.678	17.1
4	0.38	0.791	14.5	0.33	0.718	10.2
5	0.38	0.780	11.4	0.31	0.737	7.3
6	0.39	0.780	10.2	0.31	0.750	5.7
7	0.40	0.781	10.5	0.30	0.758	4.8
8	0.43	0.747	11.0	0.30	0.785	4.1
Overall			65.5			82.0

Refinement of phases determined from initial model†

Cycle	R_{av}	r	Mean phase change
1	0.37	0.792	23.2
2	0.32	0.837	9.2
3	0.30	0.858	6.3
4	0.29	0.866	4.7
5	0.28	0.873	3.8
6	0.28	0.878	3.3
7	0.27	0.881	2.7
Overall			43.4

* SIR phases were refined by symmetry averaging maps calculated with coefficients $(2F_o - F_c)\exp(i\alpha_c)$. The molecular envelopes required for symmetry averaging[13] were generated from model coordinates by enclosing each atom in a sphere of 7 Å radius. The initial model for the R1-69:14mer complex (Anderson *et al.* 1985), used to produce the envelope for 5 Å symmetry averaging (see text), contained idealized straight B-DNA. Inspection of the 5 Å map revealed that a 14mer with a slight bend would fit the density better. We therefore adjusted the positions of the two half-sites as rigid bodies using the real space refinement option of FRODO, and carried out the eight cycles of 3.2 Å refinement using an envelope from this model. Mean phase changes $(\Sigma|\phi_c - \phi_o|/N)$, for cycles subsequent to the first, are for all data present in the native data set, whether or not present in the initial SIR data set.

† Phases calculated from coordinates fit to the averaged maps from *a* were refined by averaging $F_o\exp(i\alpha_c)$ maps. The envelope was constructed from the coordinates on which the initial phases were based. $R_{av} = \Sigma|F_o - F_c|/\Sigma|F_o|$

$$r = \frac{\sum |F_o||F_c| - (\sum |F_o| \sum |F_c|)}{[\sum |F_o|^2 - (\sum |F_o|)^2] \cdot [\sum |F_o|^2 - (\sum |F_c|)^2]}$$

monomer DNA-binding surfaces is influenced by interactions between individual R1-69 subunits. The tightly structured repressor domain binds to a half-site in a precise way. Any variation in twist or bend at the central base pairs changes the spatial relationship of the two half-sites and requires compensatory variation in the protein dimer contacts, with a corresponding loss in the overall energy of the complex. We expect dimer contacts to be more extensive and less deformable in the intact repressor, and indeed alterations in central base-pair sequence have greater effects on the binding of the intact molecule than on R1-69 (ref. 2). Analogous mechanisms have been suggested on the basis of indirect evidence as a contribution to the specificity of SV40 T antigen[30].

The conformation of operator bound to a dimer of R1-69 is likely to differ from the equilibrium structure of the DNA in solution, but the base sequence will in any case affect flexibility and hence its propensity to reconfigure on binding. An even greater perturbation has been seen in the complex of *Eco*RI endonuclease with its DNA binding site, where an effective unwinding of about 25° occurs at the centre of the symmetrical sequence[31]. The unwinding induced by the endonuclease is important for appropriate alignment of protein-DNA interactions on either side.

The general notion that sequence-dependent aspects of DNA conformation can influence affinity of specific binding proteins has been discussed by Dickerson[32]. The characteristics that we see here to be necessary for a good fit, both within a half-site and between half-sites, appear to be examples of such conformational effects, but the specific 'rules' developed by Dickerson[32] may not be straightforwardly applicable[2].

Operator recognition

The tight fit between 434 repressor and DNA could not be achieved with a random 14-bp sequence. We suppose that in a non-specific complex, the repressor would be displaced by 2–3 Å outward from the DNA axis, to accommodate non-optimal major-groove interactions, and that the peptide backbone–DNA backbone hydrogen bonds would be lost. Residues such as Arg 41 (which must curl back toward a phosphate in the specific complex) and Arg 43 (which lies in the minor groove) could extend to provide compensating electrostatic stabilization. Such non-specific complexes may be intermediates in operator binding, because proteins like 434 repressor are believed to locate their sites by initial nonspecific interaction with distant sequences, followed by diffusion along the backbone. Record *et al.*[33] have postulated a similar distinction between complexes of *lac* repressor with operator and non-operator sequences.

We do not yet have a complete description of how the 434 repressor distinguishes among its six naturally-occurring operator sites. To summarize our current picture of specific repressor-operator interactions, it is useful to distinguish three zones in a 7-bp half-site. (1) The first three base pairs (ACA...), invariant in all operators (see ref. 2), interact directly with Gln 28 and Gln 29 on α3. Changes at any one of these positions decrease repressor affinity over 150-fold (R. Wharton, personal communication), and indeed no other DNA sequence can be properly complementary to the contact surface presented by the glutamine side chains. (2) The fourth and fifth base pairs both interact with Gln 33. We therefore expect effects of altered base sequence at these positions to be correlated. One of the naturally-occurring operators (O_R3) differs from the others at position 4 in one half-site, and there is considerable variation among sites at position 5. (3) The sixth and seventh base pairs form no direct contact with protein, and modulation of affinity by sequence variation at these positions occurs through effects on DNA conformation. The corresponding free energy changes are modest (1–2 kcal mol^{-1}), but natural operator sites differ in binding free energy by just such small amounts. The sequences of the six sites vary widely in this zone.

Table 3 Amino-acid residues in α-helices

	434 repressor	λ repressor
α1	1–13	9–23
α2	17–23	33–39
α3	28–36	44–52
α4	44–53	61–69

The Cα backbone of λ repressor N-terminal domain[5] was superimposed on the R1-69 backbone, optimizing coincidence of α2 and α3 (ref. 4). Assignments of helical residues were made by visual inspection of backbone conformation. Residue 15 of λ repressor closely corresponds in three-dimensional position to residue 1 of 434 repressor; 434 repressor lacks the N-terminal arm of λ repressor and the first few residues of α1. The first helix of λ repressor, as built by Pabo and Lewis, ends 'earlier' than α1 of 434 repressor (residue 23 of λ corresponds to 9 of 434). The region at the end of α1 was not well-defined in the original λ repressor map, however, and the actual Cα coincidence in α1 may be even greater than suggested by the assignments above. In the present 434 repressor model, residue 27 might also be assigned to helix 3, since its carbonyl appears to hydrogen bond to N31.

The distinction between these zones is only heuristic and it is clear that a perturbation at one position will affect the fit elsewhere. For example, the entire operator half-site must be configured correctly against the repressor monomer for any of the interactions with base pairs to be meaningful. Given such a configuration, the present description appears to explain many of the observed effects of operator mutations.

We thank Marie Drottar for laboratory assistance and preparation of crystals; R. Crouse for maintaining X-ray apparatus; J. Katz for artwork; A. Aggarwal, A. Mondragon, J. Moulai and G. Quigley for programs and advice; J. Koudelka and R. Wharton for discussion of unpublished results; C. Wolberger and D. C. Wiley for comments and suggestions. J. E. Anderson was a Burroughs Wellcome Fund Fellow of the Life Sciences Research Foundation. The work was supported by a grant from the NIH to S.C.H. and M.P.

Received 8 January; accepted 1 April 1987.

1. Ptashne, M. *A Genetic Switch* (Cell Press, Cambridge, Massachusetts, 1986).
2. Koudelka, J., Harrison, S.C. & Ptashne, M. *Nature* **326**, 886–888 (1987).
3. Wharton, R. thesis, Harvard Univ. (1985).
4. Anderson, J., Ptashne, M. & Harrison, S. C. *Nature* **316**, 596–601 (1985).
5. Pabo, C. O. & Lewis, M. *Nature* **298**, 443–447 (1982).
6. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754–758 (1981).
7. McKay, D. B. & Steitz, T. A. *Nature* **290**, 744–749 (1981).
8. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. *Nature* **317**, 782–786 (1985).
9. Pabo, C. O. & Sauer, R. *A. Rev. Biochem.* **53**, 293–321 (1984).
10. Anderson, J., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **181**, 1307–1311 (1984).
11. Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **16**, 39–45 (1973).
12. Bricogne, G. *Acta crystallogr.* **A130**, 395–405 (1974).
13. Bricogne, G. *Acta crystallogr.* **A132**, 832–847 (1976).
14. Hogle, J., Chow, M. & Filman, D. *Science* **229**, 1359 (1985).
15. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
16. Hendrickson, W. A. & Konnert, J. H. *Biomolecular Structure, Conformation, Function and Evolution* Vol. I (ed. Srinivasan, R.) 43–57 (Pergamon, Oxford, 1981).
17. Sussmann, J., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **33**, 800–804 (1977).
18. Ten Eyck, L. *Acta. crystallogr.* **A33**, 486–492 (1977).
19. Hogle, J., Maeda, A. & Harrison, S. C. *J. molec. Biol.* **191**, 625–638 (1986).
20. Wilson, I. A., Wiley, D. C. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
21. Drew, H. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2179–2183 (1981).
22. Bushman, R., Anderson, J. E., Harrison, S. C. & Ptashne, M. *Nature* **316**, 651–653 (1985).
23. Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. *Nature* **273**, 443–446 (1978).
24. Pflugrath, J. W. & Quiocho, F. A. *Nature* **314**, 257–260 (1985).
25. Dreusicke, D. & Schulz, G. *FEBS Lett.* **208**, 301–304 (1986).
26. Seeman, N. C., Rosenberg, J. N. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **173**, 804–808 (1976).
27. Baker, E. N. & Hubbard, R. E. *Progr. Biophys. molec. Biol.* **144**, 97–179 (1984).
28. Wharton, R. & Ptashne, M. *Nature* **326**, 888–891 (1987).
29. Drew, H. R. & Travers, A. A. *Cell* **37**, 491–502 (1984).
30. Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. & Tegtmeyer, P. *Cell* **44**, 719–725 (1986).
31. McClarin, J. A. *et al. Science* **234**, 1526–1541 (1986).
32. Dickerson, R. E. *J. molec. Biol.* **166**, 419–441 (1983).
33. Record, M. T., deHaseth, P. L. & Lohman, T. M. *Biochemistry* **16**, 4791–4796 (1977).
34. Crawford, J. thesis, Harvard Univ. (1977).
35. Durbin, R. *et al. Science* **232**, 1127–1132 (1986).
36. Kabsch, W. *Acta crystallogr.* A **33**, 922 (1976).

Printed in Great Britain by Turnergraphic Limited, Basingstoke, Hampshire