The complex between phage 434 repressor
DNA-binding domain and operator site O_{R3}:
structural differences between consensus
and non-consensus half-sites

David W Rodgers and Stephen C Harrison*

Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biology,
Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Background. The repressor of phage 434 binds to a
set of operator sites as a homodimer. Its relative affinities
for these sites determine the switch from lysogenic to
lytic growth. The six 434 operator sites (O_{R3}, O_{R2}, O_{R1},
O_{L1}, O_{L2} and O_{L3}) have a particularly simple organiza-
tion; all are 14 base pairs long, with a conserved 5’-
AACA sequence symmetrically placed at either end, and
a variable central six base pairs. O_{R3} is unique among
naturally-occurring 434 operator sites in that it contains
a non-consensus base pair, G-C, at the fourth position of
the otherwise invariant 5’-AACA sequence. Comparisons
among structures of the 434 repressor DNA-binding do-
main, RI–69, bound to various operator sites, allow us
to analyze differential specificity in regulatory complexes
of this kind.

Results. We have determined the structure at 2.5\AA
resolution of a complex of RI–69 with DNA containing
the O_{R3} site and compared it with previously studied
complexes of RI–69 bound to O_{R1} and O_{R2}. There are
surprisingly extensive structural differences between
the consensus and non-consensus half-sites of O_{R3} with re-
spect to their interactions with RI–69, including a shift
in the DNA backbone and a small rotation of the entire
RI–69 monomer.

Conclusions. Recognition of the base pair difference
that is critical for the 434 regulatory switch involves a
number of amino acid residues, not just the one or two
side chains in direct contact with the G-C base pair.
Moreover, the repressor imposes a somewhat altered
DNA conformation on the non-consensus half-site.

Structure 15 December 1993, 1:227–240

Key words: DNA recognition, operator, phage 434, repressor, X-ray structure

Introduction

The temperate bacteriophages such as \lambda and 434 can
switch from lysogeny to lytic growth in response to
inducing stimuli [1]. The molecular mechanism of this
switch depends on two proteins, known as repressor and Cro, on
and their relative affinities for sites in the
major operators, O_{R} and O_{L} [2]. The operator O_{R} con-
trols two distinct promoters, known as P_{R} and P_{RM}.
Promoter P_{R} governs transcription of genes, includ-
ing the one for Cro, that are important for initiating
lytic growth. Repressor turns off P_{R} by cooperatively
occupying sites O_{R1} and O_{R2}. Its presence at O_{R2} also
stimulates transcription of its own message from P_{RM}.
High concentrations of repressor lead to binding at
O_{R3} and repression of P_{RM}. The characteristics of O_{R3}
thus determine the steady-state level of repressor in a
lysogenic cell. It must have a lower affinity for repressor
than O_{R1}, so that transcription from P_{RM} is not shut
off without effective ‘shut off’ of P_{R}, but it must not have so
low an affinity that the resulting high steady-state con-
centration of repressor renders the switch insensitive
to induction.

A diagram summarizing the switch just described to-
gether with the sequences and binding characteristics
of the sites in O_{R} of phage 434 is shown in Fig. 1.

The 14 base pair sites are identical at positions 1–4 in
the left and right half-sites, except for O_{R3}, which has an
A-T \rightarrow G-C substitution at position 4 in the right
half-site (4R). Affinity measurements on model opera-
tors demonstrate that substitution at position 4R signif-
ically reduces binding of repressor without affecting
binding of Cro [3–5]. Similar experiments also show
that the sequence of the central six base pairs, which
varies from operator to operator, influences repressor
affinity, but less dramatically than any change in the
conserved outer four positions [6].

The 434 repressor binds as a dimer. Its two-fold sym-
metry corresponds to the approximate symmetry of the
operator sites. Each repressor subunit folds into two
domains. The amino-terminal domain of 69 residues,
known as RI–69, binds DNA; the carboxy-terminal do-
main mediates dimerization [7]. 434 Cro, a protein
of 71 residues, is extremely similar in sequence and
structure to RI–69 [8–10]. Structures for RI–69 and
434 Cro, alone and in complex with O_{R1}, have been
described previously, and a report of the structure for
an O_{R2}/RI–69 complex has recently appeared [11].
RI–69 and Cro form similar complexes. The ‘recog-
nition helix’ of the helix-turn-helix motif lies in the
major groove, and various side chains contact groups

*Corresponding author.

© Current Biology Ltd ISSN 0969-2126

227
at the edges of the outer four base pairs. The interface region is shown in Fig. 2a. Recognition of
the invariant ACA (positions 1-3) involves Thr27, Gln28
and Gln29, as well as supporting interactions from pe-
ripheral residues such as Gln17. Specificity at base pair
4, the principal focus of the present report, clearly in-
volves Gln33, which donates a hydrogen bond to O4
of thymine 4' in the consensus 5'CAGA sequence. But,
in the OR3 right half-site, a hydrogen bond from N4
of cytosine 4' to N6 of Gln33 should also be possible,
with rotation of the side chain by 180° about the Cα-Cβ
bond (x3). No other hydrogen-bonding interactions
appear to prevent this interchange of donor and ac-
ceptor groups. Moreover, mutation of residue 33 to ala-
nine or leucine (its identity in Cro) reduces, but does
not eliminate, the strong preference for A-T [5]. It was
therefore proposed that van der Waals contacts of the
thymine methyl group — to Cα of serine and to the
atoms of the CH2 group of Gln33 itself — might
also play a role in determining the specificity at posi-
tion 4 [12].

In order to explore our understanding of differential
specificity in DNA/protein complexes such as the ones
just described, we have determined the structure of an
OR3/R1-69 complex. The result is surprising in the
extent of the differences between the non-consensus
and consensus half-sites (Fig. 2b). The structure of
the complex for the consensus half-site (at base pairs 1-4)
is essentially identical to that seen in the OR1 and OR2
complexes. The non-consensus site shows not only a
rearrangement of the Gln33 side chain, but also a sig-
nificant shift in the DNA backbone around nucleotide
4' and a small rigid-body rotation of the entire R1-69
monomer. These shifts are also present in the struc-
ture of a complex between R1-69 and DNA with an
A-T → G-C substitution in the right half-site of OR1
[OR1(4RG), J.W. Shimon and S. Harrison, unpublished
data]. Thus, in the case of the 434 repressor, the evolu-
tion of a precisely graded differential affinity appears
to have required a remarkably distributed set of changes.
Recognition of base pair 4 involves not just Gln33 and
Ser30 but a number of residues in R1-69, not all of
which actually make contact with DNA.

Results
Overview
The sequence of the DNA fragment used for crystal-
ization of the OR3/R1-69 complex is shown in Fig.
3 together with the numbering scheme for individual
nucleotides. As with other crystals of 434 proteins com-
plexed to operator sites [10-12], the DNA fragments
stack end-to-end, pairing overhanging bases and form-
ing pseudo-continuous helices that, in this case, run
parallel to the crystallographic b axis. Bases outside
the 14 base pair (bp) operator were chosen to match
the corresponding bases in the fragment used for the
OR1/R1-69 structure. They do not reproduce the se-
quence flanking OR3 in the 434 genome.
Fig. 2. (a) Diagram showing the interface in the major groove between R1-69 and the left half-site of O₅3/R1 [13]. Residues are identified by the single-letter code. Short bars emanating from the protein main chain are peptide -NH groups. Hydrogen bonds are represented by dashed lines, and phosphates by numbers in circles along the DNA backbone. The small circle is a water molecule. Base pair numbers appear to the left of the drawing. (b) A similar diagram for the non-consensus (right) half-site of O₅3/R1-69, described in this paper. The broadly dashed curve along the DNA backbone near phosphate 4' represents the backbone position in the consensus structure [compare with part (a)]. (c) A similar diagram for the left half-site of the O₅3/Cro complex [10].
In Fig. 4a, backbone traces for the O$_3$/R1–69 and O$_1$/R1–69 structures are superimposed based on phosphorous atom positions. In the left half-site, where the sequences are the same, both the protein and DNA conformations are essentially identical to the estimated error of the coordinates. The root mean square (rms) residual on phosphorous and \( \sigma \)-carbon atoms is 0.44 \( \AA \). In the right half-site, however, the two structures superimpose less well. The differences in the right half-site are also apparent when the O$_3$ left half-site is superimposed on it, again based on phosphorous atom positions (Fig. 4b). The most marked change is in the DNA backbone at positions 3' and 4'. In the right half-site, the backbone bows out towards the protein relative to the left half-site, bringing the phosphates closer to residues in the helix 2/helix 3 turn, as well as the amino terminus of helix 3. The difference in phosphorous atom position between the left and right sides is 1.9 \( \AA \) for P4' and 1.3 \( \AA \) for P3' and P4'. Significant changes in backbone phosphorous atom positions are limited to these two nucleotides; the remaining phosphorous atoms align well (rms residual of 0.43 \( \AA \)). The rotation required to superimpose the backbone of the two halfsites, based on the remaining phosphorous atoms, is 178.9°, almost a strict dyad.

In addition to the difference in the DNA backbone, the protein monomer in the right half-site rotates relative to the left monomer, further decreasing the gap between the DNA backbone and the helix 2/helix 3 turn. Beginning from the superposition based on phosphorous atoms, an additional rotation of 5.4° is required to align the protein monomers. The shift appears to be primarily a rigid-body rotation; when optimally aligned, the right and left monomers superimpose with an rms residual of 0.51 \( \AA \). This value is comparable to that found, for example, for the superimposed left monomers of O$_3$ and O$_1$. The largest component of the rotation occurs about an axis parallel to the DNA axis, but there is a significant component of rotation about the crystallographic c axis, or roughly perpendicular to the view shown in Fig. 4. The center of rotation is near the amino terminus of helix 3, the helix that is inserted in the major groove. This fulcrum can be determined from a plot of rms distance between equivalent residues in the left and right monomers (Fig. 5). The distance between monomers has a clear minimum around residues 28 to 30, precisely those residues in helix 3 that interact with invariant base pairs 1–3 at the ends of the operator sites.

**Base pair 4R**

The interaction between the protein and a consensus A-T at position 4 has been described for the O$_1$/R1–69 complex (Fig. 2a; DW Rodgers and SC Harrison, unpublished data) [12]. A hydrogen bond links O4 of thymine 4'R with the terminal NH$_2$ of Gln33. In addition, the thymine methyl group is in van der Waals contact with the side chains of Gln29 and Ser30. Gln33 is also linked to P5' through a solvent molecule. The same interactions are made in the left, or consensus, half-site of O$_3$/R1–69 (Fig. 6a). A potential component of the discrimination against a G-C pair at position 4 might be a change in these interactions. In principle, the hydrogen bond contact to the base could still be made, now from the N4 of the cytosine to the carbonyl oxygen of Gln33, after rotation about the C5'–C6 bond. The thymine 5-methyl interaction would, of course, be lost. The details of this region in the model of the O$_3$/R1–69 right half-site are shown in Fig. 6b and a superposition of the left and right half-sites is presented in Fig. 6c.

In the vicinity of Gln33, the protein main chain has shifted away from the DNA by \( \simeq 1 \AA \) relative to its position in the left half-sites. This shift appears to be part of the overall rotation of the right monomer described above. It is accompanied by a rearrangement of the Gln33 side chain, which turns away from the base relative to the left half-site position. In this orientation, there could be no direct hydrogen bond contact with the N4 of cytosine. The gap left by the movement of the glutamine side chain is filled by a solvent molecule, which is well positioned to bridge between the glutamine side chain and N4. Thus, a previously direct interaction appears to be replaced by a solvent bridge. The distance from the OE of Gln33 to the solvent molecule is 3.2 \( \AA \), and the solvent is 3.3 \( \AA \) from the cytosine -NH$_2$ group.
Given the modest resolution and corresponding uncertainties in the map, it is possible that the glutamine side chain could have been built into density assigned to the solvent molecule, putting it closer to the position it adopts with the consensus A-T base pair. There are, however, clear indications in the 2F_o−F_c and F_c−F_e electron density maps that this alternative orientation of the side chain would be incorrect (Fig. 7). Moreover, because of the shift in the protein backbone around position 33, even if the Gln33 side chain were built as close as possible to the base, the distance to the N4 would be too great (3.7 Å) for a good hydrogen bond interaction. Thus, we conclude that the interaction between Gln33 and the base at position 4 is weakened by the G-C substitution and that no direct hydrogen bond is present.

A solvent molecule fills the void left by the rearrangement of the Gln33 side chain, but the structure shows no corresponding solvent to replace the 5-methyl lost in going from T to C at position 4' R. The gap appears instead to be taken up, at least partially, by the movement of several protein and nucleic acid groups closer
to the cytosine base. The solvent molecule replacing the terminal groups of Gln33 is actually closer to the 5 position of C4'R than the glutamine side chain, closing down the hole somewhat. The movement of the 4'R phosphate group, discussed in greater detail in the next section, narrows the gap further. Finally, the side chains of Ser30 and Gln29, as well as the main chain in this region, shift towards the DNA to help fill the void (see Fig. 8).

Another consequence of the base substitution at position 4', possibly coupled to the nature of the adjacent sequence towards the center of the operator, is a shift of the bases in the region of 4R along the helix axis (Fig. 8). This shift may result from the altered position of the DNA backbone near position 4', it may be an effect of creating the poly(dA)-poly(dT) run at the center of the operator, or it may require both. The O4(4RG)/R1–69 structure (LJW Shimon and SC Harrison, unpublished data) shows the phosphate shift at position 4R, but bases near position 4 are not displaced. O3(4RG) does not have the poly(dA)-poly(dT) central sequence of O3, and it therefore appears that the nature of the central sequence may play a large role in the shift of the bases seen in the present structure. Since these are primarily shifts along the helix axis, the alterations in base position seem not to affect the stacking interactions greatly. This visual assessment is confirmed by energetics calculations using the program CHARMM [13], which indicate that there are no major differences in stacking interactions between the left and right half-sites in this region (E Simon, personal communication).

**Phosphate 4'R**

The changes in the interaction with cytosine 4'R in O3/R1–69 are accompanied by a change in the sugar-phosphate backbone in the same region of the operator. Most prominent are the movements of phosphates 3'R and 4'R out towards the right monomer relative to their positions in the left (consensus) half-site (see Fig. 4). The shift of phosphate 4'R in particular, as well
as movement of the nearby protein, alters its interaction with the residues in the helix 2/helix 3 turn and near the amino terminus of helix 3. P3'R is located far enough round the helix that even with its movement towards the protein, there are no direct or solvent-mediated interactions with the monomer bound to the right half-site.

Details of the interaction between P4'R and the protein monomer in the left (consensus) half-site of

---

**Fig. 7.** Electron density showing the position of Gln33R. Phases for the maps came from the original molecular replacement solution (modified O91/ R1-69 model) after positional refinement as a single rigid body. The Gln33R side chain was omitted for the calculation. Positive $F_o - F_c$ density, contoured at $2\sigma$, is shown in green, and $F_o - F_c$ density at $1\sigma$ is shown in blue. The model used for phasing is shown in red and the final refined model is in yellow. Both the $F_o - F_c$ omit and $2F_o - F_c$ density clearly show the correct position of the Gln33R side chain in the final model. No bias towards this position is present in the phasing model.

**Fig. 8.** Interaction at phosphate 4'R. (a) The left half-site. The interaction with P4'R is mediated by solvent molecules. Hydrogen bonds (broken lines) are made from Thr27L and Ser30L to a water molecule, which in turn donates to the oxygen of P4'R. Thr26L interacts with another water molecule, donating in turn to the same phosphate oxygen. These interactions are part of a larger network of solvent-mediated contacts. (b) The right half-site. The shifts of P4'R and the right monomer result in a direct interaction of Thr27R and Ser30R with P4'R. The Oγ of Thr26R is oriented towards P4'R, but the distance from the phosphate oxygen is too long (3.5 Å) for assignment of a hydrogen bond. The strength of the phosphate interaction with residues 26R and 27R is probably enhanced in both half-sites, however, by the interaction of the threonine side chains with main chain NH groups. Thr26R interacts directly with the NH group of Thr27R, and although the distances are greater, the Oγ of Thr27R is close to NH groups from residues 29R, 30R and 31R at the amino terminus of helix 3. (c) Superposition of the left (open lines) and right (solid lines) half-sites.
O$_{13}$/R1–69 are shown in Fig. 8a. The distance between the phosphate group and the protein is too great for direct hydrogen bond contacts. The hydroxyl groups of the side chains of Thr24, Thr27 and Ser30 do, however, participate in a hydrogen bond network that includes an interaction with P4'/L through bound solvent molecules. The distances from one solvent molecule to the phosphate, Ser30 and Thr27 (2.6 Å, 3.0 Å and 2.7 Å, respectively) and from the other solvent molecule to the phosphate and Thr26 (2.4 Å and 2.8 Å) are all consistent with good hydrogen bond contacts to waters. The positions of the phosphate and the three protein side chains seen here are representative of those found in complexes of R1–69 with the O$_{R1}$ and O$_{R2}$ operators, which have the consensus AT pair at position 4 in both half-sites. This arrangement then seems to be the rule for the consensus sequence at positions 1–4. It should be noted, however, that the bridging solvent molecules were not assigned in either half site of the O$_{R2}$/R1–69 complex [11]. There are no identifiable features of the model that would account for their absence, and it seems most probable that they were not seen because of limited accuracy in the model rather than because of the true absence of bridging solvent molecules.

In the right half-site of O$_{R3}$/R1–69, the shift of P4'/R and the protein towards each other alters the nature of this contact (Figs 8b and 8c). The shift of P4'/R is nearly 2 Å, as noted previously, and the protein moves by nearly 1 Å for the three residues involved. An F$_{o}$–F$_{c}$ electron density map, phased with the original molecular replacement solution after rigid positional refinement only, clearly shows the phosphate shift (Fig. 9). The isomorphous model for the atoms in this phosphate group (and 3'R as well) are well within the range found for the remaining phosphate groups in the structure. The displacement of P4'/R excludes the solvent molecule present in the left half-site, and the phosphate interacts directly with the side chains of Thr27, Ser30 and possibly Thr26. The Oy atoms of Ser30 and Thr27 are well positioned for hydrogen bond formation to the phosphate oxygen, both at distances of 2.7 Å. Thr26 is farther from the phosphate, at 3.5 Å. The shift of P4'/R and changes in its interaction with the protein are also found in the structure of O$_{R1}$/R1–69 (JW Shimon and SC Harrison, unpublished data). This observation confirms the conclusions made here and establishes that the displacement of the backbone is due to the substitution of G-C for A-T rather than effects of the adjacent sequence changes in the center of the operator.

The movements of P4'/R and P3'/R are not accompanied by corresponding translations of the sugar groups at these positions (see Fig. 6c). The shifts of the phosphate groups are instead accomplished by rotations of the sugar groups and changes in backbone torsion angles. The sugars at positions 3'R, 4'R, and 5'R are all rotated with respect to the equivalent positions in the left half-site. This rotation is greatest at position 5'R. Its effect, particularly at 4'R and 5'R, is to shift C1' and the attached base towards the center of the operator, and C4'/C5' towards the end of the operator and out towards the protein. This rotation accounts for much of the shift in the phosphate positions. In general, the backbone torsion angles in this region are well within the range expected for B DNA. A particularly small 6 angle (81° rather an expected value of ~125°) at position 5'R does contribute to the P4'R shift. Small changes in other backbone torsion angles are responsible for the remaining alterations in phosphate positions.

At the 2.5 Å resolution limit of the structure, the sugar puckering angles are not well defined. Nonetheless, nearly all sugar groups outside the 3'R to 5'R region remain close to the C2' endo conformation imposed by the idealization procedure used in refinement. The sugars at positions 3'R and 4'R, however, consistently flatten during refinement after idealization. At the resolution

![Fig. 9. Electron density showing the position of phosphate 4'R. As in Fig. 7, the model used for phasing (shown in red) was the original molecular replacement solution after rigid positional refinement, and the 4'R phosphate group was omitted for the calculation. Therefore, no bias exists towards the final phosphate position. Positive F$_{o}$–F$_{c}$ density, contoured at 2.7σ is shown in green and 2F$_{o}$–F$_{c}$ density, contoured at 1σ, is in blue. Both maps demonstrate the shift of the phosphate group to its position in the final O$_{R3}$/R1–69 model (drawn in yellow). P3'/R was not omitted, but density indicating its shift in position is present.](image-url)
of the data, it is not possible to determine whether this flattening is evidence for a strained conformation.

Central bases
There are only a small number of interactions with the central six base pairs of the operator; the extensive contacts in this region are with the sugar-phosphate backbone. Yet the exact nature of the central sequence can influence the binding affinity of repressor for a particular operator [6]. This effect (sometimes called 'indirect readout' [14]) has been ascribed, at least in part, to an influence on the energetic cost of placing the central region backbone into the precise conformation required by the bound protein [6,11,12,15]. The 5 bp poly(dA)-poly(dT) sequence found in the O$_R$3 operator has been shown to be a particularly favorable central sequence [15].

Tracts of poly(dA)-poly(dT) are known [16-18] to adopt the same overwound conformation with a narrow minor groove found in all operator/R1–69 or operator/Cro structures we have determined. This conformation is thought to be stabilized by the uniform tilting of adjacent bases to maximize stacking interactions and the formation of hydrogen bonds between adjacent base pairs, known as three-centered or bifurcated hydrogen bonds. A view of the bases in the central region of the operator from the O$_R$3/R1–69 complex is presented in Fig. 10. The bases in this region adopt a very regular poly(dA)-poly(dT) type of conformation, with large propeller twist angles and a series of bifurcated hydrogen bonds. The three centered hydrogen bonds all meet length and geometry criteria proposed for this type of interaction [19], with the exception of a somewhat long bond between adenine 7'R and thymine 6'R (3.6 Å). With the inclusion of this 7'R-6'R bond, all four possible three-centered interactions are present within the 5 bp poly(dA)-poly(dT) tract. In addition, there is a bond that extends outside the poly(dA)-poly(dT) tract from the N6 of adenine 5'R to O6 of guanine 4R.

Given the structure of the central region of the O$_R$3 operator, it seems likely that the poly(dA)-poly(dT) sequence is favorable for repressor binding because it readily adopts the conformation of the bound operator fragment. The favorable central sequence of the O$_R$1 operator contains a shorter 3 bp poly(dA)-poly(dT) tract. It too has several bifurcated hydrogen bonds and shows large deviations from co-planarity of the base pairs, although these are less regular than those in O$_R$3 [12]. In contrast, O$_R$2, which binds intact repressor more weakly than either O$_R$1 or O$_R$3 (Fig. 1), has an alternating purine-pyrimidine sequence, and the bases in the central region of the operator show only small deviations from co-planarity, with no three-centered hydrogen bonds [11]. The effect of the central sequence in O$_R$2 is greater for intact repressor than for R1–69 (Fig. 1b). It is possible that intact repressor, which has a more extended dimer interface [7], may constrain the conformation of the central region of the operator even more strictly.

The only side chain of R1–69 that enters the minor groove and potentially influences central base pair specificity is Arg43. The conformations of the two Arg43 side chains are different (Fig. 11). One (Arg43L) projects directly into the groove. Its conformation is similar in complexes with O$_R$1 and O$_R$2. A water molecule, also present in the O$_R$1 and O$_R$2 complexes, intervenes between an N7 of this arginine and hydrogen-bond acceptor groups on A7'R and T7'L. The conformation of Arg43R varies more from structure to structure. In the present complex, it can hydrogen bond directly through N7 to phosphate 7'L, and two waters form a bridge between N7N2 and the conserved water described above. In the O$_R$1 and O$_R$2 complexes, Arg43R has a more extended conformation.

Comparison with Cro complex
The 434 Cro does not discriminate against a G-C substitution for A-T at position 4 of an operator site. A superposition of the O$_R$1/Cro backbone [10] on O$_R$3/R1–69 is shown in Fig. 12a. There are several regions where

Fig. 10. Central bases showing propeller twist and three-centered hydrogen bonds. Only non-Watson-Crick hydrogen bonds between strands are shown (broken lines). All four possible three-centered bonds in the 5 bp poly(dA)-poly(dT) sequence are present, although the distance for the interaction between Ade7'R and Thy6'R is rather long (3.6 Å). In addition, an interaction between N6 of Ade5'R and O6 of Gua4'R extends outside this region.
the DNA phosphorous atoms differ between the two structures, but a key deviation can be seen in the left half-site. P4'R in the O₉R₁/Cro model is shifted out towards the protein monomer in the same way that P₄'R of O₉R₃ shifts towards the protein. The distance between the phosphorous atoms is 2.0 Å, the same as the value found in the non-consensus half-site of O₉R₃/R1-69. Both half-sites of O₉R₁/Cro have this conformation, and the similarity to the O₉R₃ right half-site at position 4 is evident from the superposition. Cro makes a set of hydrogen bonds to P₄'R that resemble those made by repressor in the right half-site of O₉R₃ (Fig. 12b). The contact from Ser30 is identical. Instead of a threonine at position 27, Cro has a lysine, but a hydrogen bond from this residue to P₄'R is still present — now from the main-chain amide group instead of from the side chain. Cro has a valine instead of a threonine at position 26, and there is no interaction between P₄'R and this residue.

There is a leucine in Cro at position 33 instead of the glutamine in repressor, so a possible mechanism for...
the differential response at position 4 might be the loss of a hydrogen bond to the base pair. Reciprocal mutations at this position do not, however, reverse the preferences at position 4 for either repressor or Cro [5]. The similarities in contacts at position 4 suggest another possibility. The presence of a G-C at position 4 forces a backbone change in the repressor/operator complex, presumably at some energetic cost. Since the backbone already adopts this conformation in a complex with Cro, even with A-T at position 4, there is no equivalent energetic cost of a G-C substitution.

Discussion

The properties of the non-consensus (right) half-site of O_{R3} are critical for the phage 434 regulatory switch [2,20]. The combined interactions of repressor with the consensus and non-consensus half-sites must result in an affinity for O_{R3} that appropriately balances repression at P_{R} and P_{L} and autoregulation at P_{RM}. The structure of O_{R3}/R1–69 reported here reveals an unexpectedly extensive set of changes in the non consensus half-site with respect to the consensus half-site and the O_{R1} complex. These changes include a displacement of phosphates 4' and 3', a small rigid-body rotation of the R1–69 monomer, a rearrangement of the Gln33 side chain, and a shift in the stacked bases of nucleotides 4' and 5'. The structure of O_{R1}(4RG)/R1–69 (IJW Shimon and SC Harrison, unpublished data) shows many of the same differences — in particular, the phosphate displacement and a rigid-body motion of the monomer. Thus, changes in one or two side chain to base contacts do not alone account for recognition of the substitution at position 4. Interactions are spatially distributed across the protein/DNA interface. The diagram in Fig. 13 summarizes R1–69/operator contacts seen in the two structures. This representation emphasizes only the local differences, many of which result from the larger-scale adjustments just mentioned — rearrangement of the DNA backbone and rotation of the R1–69 monomer.

In the consensus half-site structure, there is an elaborate hydrogen-bond network that links Thr26, Thr27, Ser30 and the water molecules bridging to phosphate 4'. This arrangement appears to be reinforced by van der Waals contacts between the thymine 4' methyl group and various atoms on the protein — Cβ of Gln29, Oγ of Ser30, and the CONH₂ group of Gln33. One consequence is that phosphate 4' is 'pushed' away from the helix 2/helix 3 turn and into the minor groove, contributing to its extreme narrowing in this region. In the non-consensus half-site, the phosphate returns to a more conventional B-form position. It reorganizes the hydrogen bond network by inserting an oxygen between Thr27 and Ser30 and by expelling the bridging water molecule. The void left by the absence of a pyrimidine methyl group closes down as well, and C5 of cytosine 4' is in van der Waals contact with Oγ of Ser30 and with the water that bridges between N4 of the cytosine and Oε of Gln33.

It is noteworthy that ordered water molecules sometimes replace protein or DNA hydrogen-bonding groups in regions where the two half-sites are different. For example, the water that intervenes between O2P4' and Oγ of Thr27 and Ser30 in the left (consensus) half-site is effectively replaced by the phosphate oxygen itself in the right half-site. Conversely, in the right half-site, a water occupies the position that Gln33 Nε takes in the left half-site. In general, waters decorate the crevices at the edges of the protein/DNA interface, because the concavity of the major groove presents an awkward shape for precise protein complementarity. Important solvent-mediated contacts to
DNA have been observed in other structures, the trp repressor/operator complex [14] being a particularly striking example.

The methyl group of thymine 4' is actually at the center of the set of contacts shown in Fig. 13, with the hydrogen bond network just described on one side and Gln33 on the other. Rigid-body rotation of the R1–69 monomer couples events around phosphate 4' with the interactions of residue 33, and several lines of evidence point to the thymine 4' methyl group as a critical determinant of the coupling. First, introduction of G5MeC into a model operator at base pair 4 has only a small effect on repressor binding — about two orders of magnitude less than the effect of G·C itself [5]. This striking result is probably the most important indication that the methyl group is actually the key feature of the A·T base pair at position 4 and that its removal, rather than substitution of a 4-amino for a 4-carbonyl on the pyrimidine, is the critical consequence of the A·T → G·C change in O2·3. Second, when Gln33 extends to form a hydrogen bond with O4 of thymine 4', it is also in van der Waals contact with the thymine methyl group. Cooperativity of the two interactions may therefore stabilize the extended conformation of the side chain. Substitution of a cytosine at position 4 will still allow a hydrogen bond, with a 180° rotation about χ3 of Gln33, but this bond can form only when the overall rigid-body position of R1–69 is the one seen in the consensus half-site complex. If removal of the methyl group destabilizes the hydrogen bond, it may also reduce the cost of the overall rotation of the monomer and thereby facilitate the rearrangement of phosphate 4'. Third, Cro (which is indifferent to an A·T → G·C substitution at base pair 4) has a much reduced hydrogen-bond network at the helix 2/helix 3 turn and a strong direct interaction with phosphate 4' [10]. It also lacks a hydrogen-bonding side chain at 33, but restoration of a glutamine at this position does not introduce a preference for A·T [5]. Thus, there is again a correlation between features on either side of the methyl group.

Why, despite the apparent tidiness of the rearranged interface, as illustrated in Figs 2b and 13, is substitution of G·C at base pair 4 unfavorable for repressor binding? The principal reason is likely to be that formation of proper contacts in the non-consensus half-site requires a small but significant rigid-body rotation of the repressor DNA-binding domain with respect to its consensus half-site partner and hence some perturbation of the dimer interface. Indeed, were it not for this strain on the dimer interface, the interactions in the non-consensus half-site of O2·3/R1–69 might well be more stable than those in the consensus half-site, because the gain of two strong hydrogen bonds to phosphate 4' probably more than compensates for loss of the direct hydrogen bond between Gln33 and base 4'. In addition, the free energy cost of a G·C substitution at base pair 4 is somewhat dependent on the identity of base pairs 5 and 6, and we have suggested earlier that the observed shifts in bases 4' and 5' may account for this effect. Thus, a large number of contributions, favorable and unfavorable, and involving many residues in R1–69, appear to determine the net difference between binding to O2·1 and binding to O2·3 — that is, to the balance between repression at P_R and autoregulation at P_RM.

Does the correlation of a backbone conformational change (the shift in position of phosphate 4') with a base pair substitution (A·T → G·C) mean that we are dealing here with an example of 'indirect readout'? That is, does 434 repressor recognize base pair 4 'indirectly' by sensing the backbone shift? We suggest that the answer is 'no' because we believe that the equilibrium position of phosphate 4' is actually dictated by interaction with the helix 2/helix 3 turn, which itself depends for its position on the presence or absence of a methyl group on base 4' and on the presence or absence of a hydrogen bond to this base from Gln33. In other words, the correlation between DNA backbone conformation and base pair identity requires the presence of repressor. By contrast, recognition of the central base pairs does appear to be 'indirect', to the extent that it depends on the way the free energy required to impose the observed structure varies with base sequence [11].

**Biological implications**

The structure described here is one of a series of complexes between R1–69, the 434 repressor DNA-binding domain, and various synthetic operator sites. It is the only such set of relatively high-resolution structures currently available, and our purpose in studying it is to understand how structural changes correlate with the sorts of affinity differences that are important for biological regulation.

The 434 repressor, which binds to DNA as a dimer, blocks lytic growth of 434 bacteriophage by binding cooperatively to two operator sites, O_R1 and O_R2. Binding to O_R3 blocks its own expression, and the affinity of repressor for O_R3 must therefore be lower than its affinity for O_R1. The reduced affinity is due to the substitution of G·C for A·T at position 4 in the right half-site of O_R3. Our work shows that the lack of the thymine methyl group causes a surprisingly large number of changes in protein/DNA contacts: the right half-site R1–69 monomer rotates slightly with respect to the operator, and the DNA backbone shifts in the vicinity of the altered base pair. We believe that this extensive rearrangement occurs because anything less would be substantially more destabilizing.
One suggestion from the present result is that small (1–3 kcal mol⁻¹), biologically 'useful' affinity differences in DNA recognition — such as those in the temperate phage genetic switch — may often require a distributed set of changes, in order to compensate for the relatively drastic local effect of a base pair substitution. Indeed, structures of a large number of protein/DNA complexes — beginning with the 14mer/R1–69 structure reported by Anderson et al. [21] — have shown that specific recognition is a concerted property of the entire protein/DNA interface. Conversely, in evolving an altered specificity — such as the relationship between 434 Cro and repressor — a simple mutation in one or two amino acid residues may be too crude. Interchanges of local sequence, such as the 434 ‘helix swap’, can reproduce qualitative specificity changes in vitro [22], but it is much harder to ascribe the quantitative affinity differences between two proteins to just one or a few side chains that interact with the relevant bases.

### Materials and methods

**Crystallization**

R1–69 was overproduced and isolated as described previously [7], and DNA fragments containing the O₉₃ operator sequence were purchased from Pharmacia. Crystals were grown from hanging drops at 4°C. The protein and annealed DNA were pre-mixed at a ratio of 1:1 to a concentration of 0.55 mM and added to an equal volume of well solution containing 15–17% (w/v) PEG 3000, 80 mM NaCl, 100 mM MgCl₂, 2 mM spermine, and 100 mM MES pH 5.5. The initial volume of the drops was 6 µl.

The crystals grow in about 2 weeks as flattened rods with maximum dimensions of 0.5 mm × 0.2 mm × 0.05 mm. The space group is P₂₁₂₁, with cell dimensions of a = 150.9 Å, b = 64.5 Å, c = 27.8 Å, and the crystals diffract to beyond 2.5 Å in the direction of the DNA axis (b cell axis) and 2.6 Å in perpendicular directions.

**Data collection**

Data were collected at 4°C on a Siemens area detector with CuKα radiation produced by an Enraf-Nonius GX-13 rotating anode generator equipped with focusing optics. The BUDDHA suite of programs [23,24] was used to reduce data from 11 crystals; the data were then merged with programs from the CCP4 package [SERC (UK) Collaborative Computing Project No. 4, Daresbury Laboratory, UK, 1979]. An analysis of the data is presented in Table 1.

**Structure determination and refinement**

The packing arrangement in these crystals deviates only slightly from the known [12] packing in crystals of the O₉₃/R1–69 complex. Beginning with the O₉₃/R1–69 model, the nucleic acid was converted manually to the O₉₃ sequence, and the modified complex placed into the O₉₃ cell. The initial R-factor was 0.527 with all data from 20 Å to 8 Å resolution. This model was then refined as a single rigid body with the program CORELS [25], reducing the R-factor to 0.323. The complex shifted 1.7 Å along the diagonal between the a and b cell axes with only a slight rotation (< 0.2°). The resolution was then extended to 6 Å, and then to 3.5 Å, while the model was broken up into progressively smaller rigid groups. Final rounds of refinement with CORELS were done at 3 Å with the DNA broken up into nucleosides and phosphate groups and the protein broken up into individual residues. Rotation about the glycosidic bond as well as main-chain and side-chain torsion angles were allowed to refine with restraints at this resolution. The R-factor after CORELS refinement was 0.285 on all data from 15 Å to 3.0 Å.

### Table 1. Data collection statistics.

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Shell (%)</th>
<th>Rshell</th>
<th>Sphere (%)</th>
<th>Rsphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.40</td>
<td>71.3</td>
<td>0.057</td>
<td>71.3</td>
<td>0.057</td>
</tr>
<tr>
<td>5.42</td>
<td>92.7</td>
<td>0.068</td>
<td>84.0</td>
<td>0.061</td>
</tr>
<tr>
<td>4.38</td>
<td>94.7</td>
<td>0.073</td>
<td>88.6</td>
<td>0.066</td>
</tr>
<tr>
<td>3.91</td>
<td>92.8</td>
<td>0.077</td>
<td>89.9</td>
<td>0.070</td>
</tr>
<tr>
<td>3.51</td>
<td>90.1</td>
<td>0.087</td>
<td>89.9</td>
<td>0.074</td>
</tr>
<tr>
<td>3.21</td>
<td>88.8</td>
<td>0.090</td>
<td>89.7</td>
<td>0.078</td>
</tr>
<tr>
<td>2.98</td>
<td>86.8</td>
<td>0.100</td>
<td>89.1</td>
<td>0.080</td>
</tr>
<tr>
<td>2.79</td>
<td>84.2</td>
<td>0.134</td>
<td>88.3</td>
<td>0.081</td>
</tr>
<tr>
<td>2.63</td>
<td>77.4</td>
<td>0.154</td>
<td>86.6</td>
<td>0.083</td>
</tr>
<tr>
<td>2.50</td>
<td>66.3</td>
<td>0.178</td>
<td>83.8</td>
<td>0.083</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>0.083</td>
<td></td>
<td>0.083</td>
</tr>
</tbody>
</table>

R value for symmetry related reflections is defined as

\[ R_{sym} = \frac{\sum_{ij}||l_i|| - <l>|\sum_{ij}||l_i||, \text{ where } l_i \text{ is the intensity of an individual measurement and } <l> \text{ is the mean value for all measurements for each independent reflection.} \]

Values are given for the indicated resolution shell (Rshell) and for all data to the indicated resolution (Rsphere). Shell is the percentage of possible reflections observed in the resolution shell. Sphere is the percentage of possible reflections observed to the indicated resolution. Data were collected from 11 native crystals. A total of 20076 observations were used for 8311 independent reflections, giving an average redundancy of 2.4.

### Table 2. Final model statistics.

<table>
<thead>
<tr>
<th>Number</th>
<th>Rms final model</th>
<th>Target rms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond length deviations</td>
<td>1888</td>
<td>0.01 Å</td>
</tr>
<tr>
<td>Bond angle deviations</td>
<td>2712</td>
<td>1.7°</td>
</tr>
<tr>
<td>Torsion angle deviations</td>
<td>1086</td>
<td>18.2°</td>
</tr>
<tr>
<td>Trigonal atom non-planarity</td>
<td>28</td>
<td>0.01 Å</td>
</tr>
<tr>
<td>Planar groups</td>
<td>176</td>
<td>0.02 Å</td>
</tr>
<tr>
<td>Bad contacts</td>
<td>42</td>
<td>0.11 Å</td>
</tr>
<tr>
<td>Isotropic thermal factors</td>
<td>1822</td>
<td>3.0 Å²</td>
</tr>
</tbody>
</table>

*Values based on observed reflections with amplitude > 1.5σ.

This intermediate model was manually rebuilt using the program FRODO [26], and the orientation of the complex about the pseudo-dyad of the operator was checked as described previously [10]. Refinement of the rebuilt model was then resumed.
with the TNT suite of programs [27]. The resolution cutoff was extended in four steps to 2.5 Å, the limit of the data. The R factor of the model on all data from 6 Å to 2.5 Å was 0.281. At that point, restrained individual isotropic temperature factors were refined followed by cycles of rebuilding and positional refinement, resulting in an R factor for the model of 0.248. Subsequent refinement of the model was carried out against data >1.5σ. The modest anisotropy in the diffraction limit of the crystals was accommodated by including an overall anisotropic scaling temperature factor with additional cycles of positional refinement and manual adjustment of the model. This decreased the R-factor to 0.213. A total of 40 solvent molecules were gradually added with further rounds of refinement and a change in the low resolution limit of included data to 5.0 Å. Prior to the final cycles of refinement, the DNA was idealized using CORELS primarily to place the sugars in C2'-endo conformation. The final R factor of the model is 0.187 against data >1.5σ from 5.0 Å to 2.5 Å. Statistics on the final model are given in Table 2. From a plot of R factor versus resolution (Fig. 14) the rms coordinate error is estimated to be about 0.5 Å.

Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory.

Acknowledgments: We thank Marie Drott for technical assistance, Anil Aggarwal for advice, Gerald Koudelka for communicating results prior to publication, and Mark Ptashe for continued interest and collaboration. The work was supported in part by NIH Grant GM-29109 (to M Ptashe) and by an American Cancer Society Postdoctoral Fellowship (to DWR).

References
