Reprinted from Nature, Vol. 326, No. 6116, pp. 886-888, 30 April 1987

Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro

Gerald B. Koudelka, Stephen C. Harrison & Mark Ptashne

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

The repressor of phage 434 binds to six operator sites on the phage chromosome[1–3]. A comparison of the sequences of these 14-base-pair (bp) operator sites[3] reveals a striking pattern: at five of the six sites, the symmetrically arrayed outer eight base pairs (four in each half-site) are identical and the remaining site differs at only one position (Fig. 1*b*). In contrast, the sequences of the inner four base pairs are highly variable. Crystallographic analysis of the repressor–operator complex[4] shows that at each half-site, the 'recognition α-helix' of the repressor is positioned in the major groove such that it could contact the outermost five base pairs, but not the innermost two (Fig. 1*a*). We show in this paper that the sequence of the central base pairs of the operator (two in each half-site) have a significant role in determining operator affinity for repressor, despite the evidence presented here and in the accompanying paper[4] that these base pairs are not contacted by repressor. We also show that these central base pairs influence operator affinity for Cro, a second gene regulatory protein encoded by phage 434 (ref. 5). We discuss the likely structural basis for this evidently indirect, but sequence-dependent, effect of the central base pairs of the operator on its affinity for the two regulatory proteins.

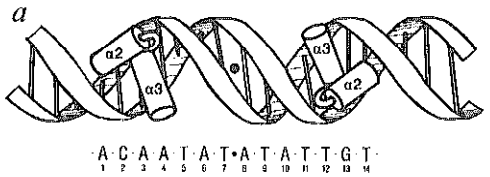
For the experiments presented here we synthesized a series of mutant 434 operators. Each of these operators differs from the twofold symmetric reference '14mer' at a single position, or at two symmetrically related positions in the central four base pairs (Fig. 1*b*). We studied the affinities of these operators *in vitro* for three proteins: 434 repressor, the amino-terminal DNA-binding domain of repressor (R1-69) and 434 Cro.

Table 1 shows that changing positions 7 and 8 (the two central base pairs) together from T·A-A·T to A·T-T·A had no effect on the affinity of the operator for repressor. In contrast, a coordinated change to either G·C-C·G or C·G-G·C decreased the affinity of the operator for repressor 50-fold. Similar results were obtained for operators that had changes at positions 6 and 9: a double substitution of A·T-T·A with T·A-A·T had no effect on the affinity of the operator for repressor, whereas both double changes, A·T-T·A → G·C-C·G and A·T-T·A → C·G-G·C at these positions decreased binding fivefold.

A single (asymmetric) T·A → G·C change at position 7 had a smaller (six- to sevenfold) negative effect on the affinity of the operator for repressor than did the double T·A-A·T → G·C-C·G changes at positions 7 and 8 (Table 1). A single T·A → A·T change at position 7 had a small (threefold) positive effect on the affinity of the operator for repressor (Table 1).

Substituting the central base pairs also affects the operator's affinity for both Cro and R1-69 (Table 2). Comparing Tables 1 and 2 reveals that the patterns of the affinity changes observed in response to the various substitutions were similar for all three proteins. Note also that any given base-pair substitution had a larger effect on the operator's affinity for the intact repressor than for either R1-69 or Cro.

Inosine, like guanine, base pairs with cytosine, but, like adenine, it lacks an N2 amino group that could protrude into the minor groove. Replacing the T·A-A·T base pairs at positions 7 and 8 with I · C base pairs has no effect on the operator's affinity for repressor, R1-69, or Cro (Tables 1 and 2).



b 434 Operator sequences

Position:	1 2 3 4 5 6 7 8 9 10 11 12 13 14
Naturally occurring operators	
O_R1	A C A A G A A•A G T T G T
O_R2	A C A A G A T•A C A T T G T
O_R3	A C A A G A A•A A A Ⓒ T G T
O_L1	A C A A G G A•A G A T T G T
O_L2	A C A A T A A•A T A T T G T
O_L3	A C A A T G G•A G T T T G T
Synthetic operators	
14mer	A C A A T A T•A T A T T G T
6T	A C A A T [T] T•A [A] A T T G T
6C	A C A A T [C] T•A [G] A T T G T
6G	A C A A T [G] T•A [C] A T T G T
6G/2	A C A A T [G] T•A T A T T G T
7A	A C A A T A [A•T] T A T T G T
7A/2	A C A A T A [A]•A T A T T G T
7C	A C A A T A [C•G] T A T T G T
7G	A C A A T A [G•C] T A T T G T
7G/2	A C A A T A [G]•A T A T T G T
7I	A C A A T A [I•C] T A T T G T

Fig. 1 *a*, Schematic representation of the 434 repressor–operator complex. The helix–turn–helix structures of the dimer are shown in the configuration they have when bound to the synthetic operator[4]. Below, sequence of the operator illustrating the position of each operator base pair with respect to the helix–turn–helix structure. *b*, Sequences of the naturally occurring and synthetic 434 operators along one DNA strand. ●, Centres of symmetry of the operators. The circled base in the sequence of O_R3 denotes the only base in the symmetrically arrayed outer eight bases of the naturally occurring operators that is not completely conserved. The bases in the synthetic operators that differ from the reference 14mer sequence are boxed.

Does repressor come into contact with the central base pairs? An examination of Fig. 1*a* suggests that a protein–DNA contact might be made on either the front side of the operator in the minor groove or on the back side in the major groove. The repressor of phage λ exploits the latter strategy, using a flexible 'arm' to make a sequence-specific contact to the major groove functional groups of the bases near the centre of its operator[6,7]; 434 repressor, however, does not have an amino-terminal arm[4,8]. The structural analysis of the repressor–operator complex[4] indicates that Arg 43 could approach bases in the minor groove in the front of the operator. To test whether this residue might allow 434 repressor to discriminate between A·T and G·C base pairs at positions 6–9, we changed Arg 43 to Ala by directed mutagenesis and studied the binding *in vitro*, of the mutant protein to the reference and mutant operators. This protein, like wild-type repressor, bound more tightly to the reference operator than to operators bearing G·C substitutions at positions 6, 7 and 7 and 8 (Table 3), although its affinity for the reference operator was decreased about 500-fold. Because the mutant protein bound to operator only at concentrations near its non-specific binding affinity, a precise value of its discrimination

Table 1 Affinity of 434 repressor for synthetic operators

Operator	Relative affinity
14mer	1
6T	1.5
6C	5
6G	5
6G/2	2.5
7A	1
7A/2	0.3
7C	50
7G	50
7G/2	7
7I	1

For nomenclature used for operators, see Fig. 1. Binding affinities are expressed as the concentration of repressor monomers needed to occupy half-maximally each operator in a DNase I protection experiment. These values have been normalized to the amount of repressor needed to half-maximally occupy the reference operator. This value was determined in an experiment performed in parallel. Under the conditions used in these experiments 1 corresponds to 2×10^{-8} M.

Methods. The operators were synthesized on an Applied Biosystems model 380A DNA synthesizer and were purified by gel electrophoresis. The isolated single strands were annealed and ligated [15] into the unique *Sal*I site in pUC18 (ref. 16). Except for the 7I operator, the ligation products were introduced into *Escherichia coli* strain JM101 (ref. 17). Plasmid DNA of resulting ampicillin-resistant colonies was screened for presence of the operator by restriction-enzyme analysis. The sequence of the operator was confirmed by DNA sequencing[18]. The operator-containing *Eco*RI-*Hin*dIII fragment, 3'-end-labelled at the *Eco*RI end, was purified by gel electrophoresis[15]. The 7I operator-containing *Eco*RI-*Hin*dIII fragment was obtained directly from the ligation products, without prior transformation. DNase I protection experiments were performed essentially as in ref. 19. The DNase I protection assay conditions were: 10 mM Tris, *p*H 7.8, 50 mM KCl, 1 mM $MgCl_2$, 2.5 $\mu g\ ml^{-1}$ sonicated chick blood DNA, 100 $\mu g\ ml^{-1}$ bovine serum albumin, 1 nM operator DNA and repressor at concentrations which varied by increments of 1.5-fold within each titration. Following incubation at 25 °C for 15 min, sufficient DNaseI was added to give, on average, one cleavage per DNA molecule in 5 min of further incubation.

Table 2 Affinities of 434 R1-69 and Cro for synthetic operators

Operator	Relative affinity	
	R1-69	Cro
14mer	1	1
6T	1.3	0.6
6C	3	2.7
6G	3	2.7
6G/2	1.8	ND
7A	1.3	1.5
7A/2	0.6	0.6
7C	15	17
7G	16	17
7G/2	4	3.2
7I	1	1

ND, not determined. The binding affinities of the proteins are expressed as in Table 1. Under the conditions of these experiments (see Table 1 legend), the concentrations of proteins required to occupy half-maximally the reference operator were: for R1-69, 5×10^{-7} M; for Cro, 2×10^{-7} M.

Table 3 Affinity of repressor(Arg 43 → Ala) for synthetic operators

Operator	Relative affinity
14mer	1
6G/2	2.5
7G	>5
7G/2	>5

Binding affinities of the 434 repressor(Arg 43 → Ala) are expressed as in Table 1. The conditions of these experiments were the same as described in legend to Table 1 except that the KCl concentration was 10 mM. Site-directed mutagenesis of 434 repressor was performed essentially as described in ref. 20. The mutant protein, bearing the Arg43 → Ala substitution, was purified following ref. 21. The concentration of this protein needed to occupy the reference operator half-maximally was 1.2×10^{-5} M.

between A·T and G·C base pairs at positions 7 and 8 could not be determined, but it is at least fivefold. The mutant protein discriminates between G·C and A·T at position 6 as well as the wild-type does (2.5-fold). The mutant protein also distinguishes between the reference operator and other higher affinity operators as well as wild-type repressor does (data not shown).

Thus, two lines of evidence suggest that the central four base pairs in the 434 operator affect its affinity for repressor by influencing DNA conformation and not by interacting directly with the protein. First, operators that have either T·A-A·T or A·T-T·A base pairs at the symmetric operator positions 7 and 8 or at 6 and 9 bind equally well to repressor, whereas operators that have either G·C or C·G base pairs at these positions bind repressor equally poorly. We know of no other set of operator mutations where the pattern A·T = T·A > G·C = C·G has been observed. For example, this pattern of changes in the affinity of the operator for repressor differs from that observed when base substitutions are made at the outer five base pairs in each half-site. Repressor prefers one base at those positions, binding poorly to operators that have any of the three mutant bases (unpublished results). The structure of the R1-69 operator complex, described in the accompanying paper, shows that all the outer position bases can be contacted directly by side chains of the recognition α-helix[4]. Second, the crystallographic investigation[4] shows that there is only one amino acid of repressor, centred just above the minor groove, that could contact the central four bases pairs; mutating this residue, Arg 43, has no effect on the ability of the protein to distinguish between A·T and G·C base pairs at the centre of the operator.

The six amino-acid residues at the carboxy terminus of R1-69 (residues 64-69) are unresolved in the structure of the protein-DNA complex, but the model suggests that no contacts between these residues and the central bases can be made[4]. Residues 64-69 cannot contact these bases in the minor groove because access is blocked by residues 40-44. Access to the central bases in the major groove is possible on the back side of the operator. However, model building suggests that even in an extended conformation, residues 64-69 cannot reasonably reach around DNA to make contact with the central bases in the major groove. In experiments not shown, we found that an intact mutant repressor, bearing alanine in the place of arginine at position 69, bound the various operators listed in Fig. 1*b* identically to wild-type repressor. Moreover, inosine, which differs from guanine only in the minor groove, behaves like adenine, not guanine, when placed in the centre of the operator (Tables 1 and 2). This result makes it highly unlikely that a contact in the major groove is responsible for the observed effects. Finally, consistent with the idea that contact is made in neither the minor nor the major groove near the centre of the operator, repressor does not protect from methylation either adenine or guanine bases that are placed in various combinations at the central positions and methylation of these bases does not block repressor binding (ref. 2 and unpublished results).

Although we do not yet know the structure of the Cro-DNA complex, the experiments of Wharton *et al.*[2] and the high degree of amino-acid sequence homology between Cro and repressor[8] suggest that it is very similar to the repressor-DNA complex. Hence, the central base pairs might also affect operator affinity

for Cro by influencing DNA conformation. We note that the amino-acid sequence of Cro differs from that of repressor at all positions homologous to repressor residues 64–69 (ref. 8). This difference suggests that these amino acids are not responsible for the observed effects on operator affinity for either Cro or repressor.

The operator in the 434 R1–69–DNA complex is overwound and slightly bent near its centre, and the minor groove of this region is compressed[4]. Evidently, G·C base pairs at positions 6–9 adversely affect binding of the protein by making the overwinding and bending of the DNA more difficult. This idea is consistent with the proposals of Drew *et al.*[9,10], who argued from the analysis of nuclease digestion experiments that the minor groove of A+T rich DNA is more readily overwound and compressed than is that of G+C-rich DNA. Drew *et al.* also concluded that minor grooves of poly(A) tracts are particularly susceptible to compression. This susceptibility could explain why an operator with the single T·A→A·T change at position 7, which creates a sequence of three A residues, has an increased affinity for protein. Application of Calladine's rules[11,12], which have been used to predict DNA structure, cannot account for the effect of G·C substitution at position 6–9 on the affinity of the operator for the proteins. In their simplest form, these rules do not distinguish between the purines; hence they would predict that operators containing G·C bases at positions 6–9 should have the same DNA structure (and by hypothesis, the same affinity) as operators containing A·T bases at these positions.

Removal of the N2 amino group of guanine, by substituting inosine, eliminates the deleterious effect of this base in our experiments. We cannot account for the negative effect of guanine by simply assuming that this N2 amino group clashes with the ring of an adjacent purine on the opposite strand: in the operator 7G/2 there is no purine positioned to make such a clash; nevertheless this operator binds the proteins less well than does the reference 14mer. We therefore imagine that the negative effect of G·C base pairs is the result, at least in part, of the base-pair hydrogen bond the N2 amino group forms with cytosine. The extra hydrogen bond could make the G·C base pair less deformable and therefore increase the energy required to form the altered DNA structure observed in the R1-69–DNA complex.

The DNA at operator positions 6–9 is configured to bring the half-sites of the operator into proper alignment with the protein, thus allowing each monomer of the bound dimer to make optimal contacts within each operator half-site[4]. Although the crystal structure reveals interactions between the two DNA-bound amino-terminal domains, biochemical experiments show that the intact repressor dimer is held together primarily by interactions between the two carboxy-terminal domains[13]. We therefore suppose that a dimer of two amino-terminal domains is more flexible than one of two intact monomers. This flexibility could explain why substituting G·C for A·T base pairs at positions 6–9 affects operator affinity for R1-69 less than its affinity for intact repressor.

The 434 repressor and Cro bind to the six naturally occurring operator sites with different orders of affinity[2,3] and these differences are crucial to the life cycle of the phage[14]. Inspection of Fig. 1*b* reveals that two sites, O_R2 and O_L1, differ only in the central base pairs. Repressor and Cro both bind more tightly to O_L1 than to O_R2 (ref. 3) and evidently this discrimination is determined entirely by non-contacted bases. Although we do not have a complete description of how these proteins differentiate among the six naturally occurring operators, we believe that the indirect effect of the identity of the central base pairs has an important function in distinguishing the sites.

We thank members of the Ptashne and Harrison laboratories, especially Robin Wharton, Ed Giniger, Cynthia Wolberger and John Anderson for materials, discussions and for providing data in advance of publication. G.B.K. is supported by a Fellowship of the Damon Runyon–Walter Winchell Cancer Fund. This work was supported by a grant from NIH to M.P. and S.C.H.

Received 8 January; accepted 1 April 1987.

1. Johnson, A. D. *et al. Nature* **294**, 217–223 (1981).
2. Wharton, R. P., Brown, E. L. & Ptashne, M. *Cell* **38**, 361–369 (1984).
3. Wharton, R. P. thesis, Harvard Univ. (1985).
4. Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 888–891 (1987).
5. Grosschedl, R. & Schwarz, E. *Nucleic Acids Res.* **6**, 867–881 (1979).
6. Eliason, J. L., Weiss, M. A. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2339–2343 (1984).
7. Pabo, C. O., Krovatin, W., Jeffrey, A. & Sauer, R. T. *Nature* **298**, 441–443 (1982).
8. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. *Nature* **298**, 447–451 (1982).
9. Drew, H. R. & Travers, A. A. *J. molec. Biol.* **186**, 773–790 (1985).
10. Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191**, 659–675 (1986).
11. Calladine, C. R. *J. molec. Biol.* 343–352 (1982).
12. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257**, 14686–14707 (1982).
13. Anderson, J. E. thesis, Harvard Univ. (1984).
14. Ptashne, M. *A Genetic Switch* (Blackwell, Palo Alto, 1986).
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
16. Yannisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103–119 (1985).
17. Messing, J. *Meth. Enzym.* **101**, 29–79 (1983).
18. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
19. Johnson, A. D., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5061–5065 (1979).
20. Zoller, M. J. & Smith, M. *Nucleic Acids Res.* **10**, 6487–6500 (1982).
21. Anderson, J. A., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307–1311 (1984).

Printed in Great Britain by Turnergraphic Limited, Basingstoke, Hampshire