Biophysical Chemistry 29 (1988) 31–37
Elsevier

BPC 01211

Recognition of DNA sequences by the repressor of bacteriophage 434

S.C. Harrison [a,b], J.E. Anderson [a,*], G.B. Koudelka [a], A. Mondragon [a], S. Subbiah [a], R.P. Wharton [a,**], C. Wolberger [a] and M. Ptashne [a]

[a] *Department of Biochemistry and Molecular Biology, Harvard University and* [b] *Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, MA 02138, U.S.A.*

Accepted 15 October 1987

DNA sequence recognition; DNA-protein binding; X-ray crystallography; Repressor-operator complex; Transcriptional regulation

The structure of a complex between the DNA-binding domain of phage 434 repressor and a 14 base-pair synthetic DNA operator reveals the molecular interactions important for sequence-specific recognition. A set of contacts with DNA backbone, notably involving hydrogen bonds between peptide -NH groups and DNA phosphates, position the repressor and fix the DNA configuration. Direct interactions between amino acid side chains and DNA bases involve nonpolar van der Waals contacts as well as hydrogen bonds. The structures of the repressor domain and of the 434 cro protein are extremely similar. There appear to be no major conformational changes in the proteins when they bind to DNA.

1. Introduction

A cascade of molecular 'switches' control the choice between lysogeny and lytic growth in temperate bacteriophages such as λ, 434 and P22 [1]. Elements of the central part of this cascade are shown in the diagram of fig. 1. The tripartite rightward operator, O_R, overlaps both the promotor transcripts required for lytic growth (P_R) and the promotor for maintenance of the lysogenic state (P_{RM}). In a lysogen, repressor is normally bound at sites O_R1 and O_R2, blocking transcription from P_R. The O_R2-bound molecule activates transcription from P_{RM}. When excess repressor is present, O_R3 is occupied and transcription at P_{RM} ceases. Inactivation of repressor, following ultraviolet irradiation, for example, leads to initiation of transcription at P_R with subsequent production of Cro (the product of the *cro* gene). Cro binds tightly to site O_R3, preventing further repressor synthesis: the transition to lytic growth is irre-

Correspondence address: S.C. Harrison, Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, MA 02138, U.S.A.

* Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, Long Island, NY 11724, U.S.A.

** Present address: Department of Neurobiology and Behavior, Columbia University and Howard Hughes Medical Institute, 722 West 168th Street, New York, NY 10032, U.S.A.

Table 1

Sequences of some 434 operator sites and corresponding relative dissociation constants for 434 repressor, R1-69 and 434 Cro

For operator sequences, see refs. 11 and 13. For affinities, see refs. 2 and 3. The 14-mer is a symmetrical sequence used for the synthetic operator in the crystal structure described in ref. 5. Relative dissociation constants are expressed as the concentration of repressor required to half-saturate a site, normalized to the concentration required to half-saturate O_R1 (repressor, R1–69) or O_R3 (Cro).

		Dissociation constant (relative)		
		Repressor	R1–69	Cro
O_R1	ACAAGAAAGTTTGT	1	1	2
O_R2	ACAAGATACATTGT		5	4
O_R3	ACAAGAAAAACTGT	12	5	1
14-mer	ACAATATATATTGT	40	9	8

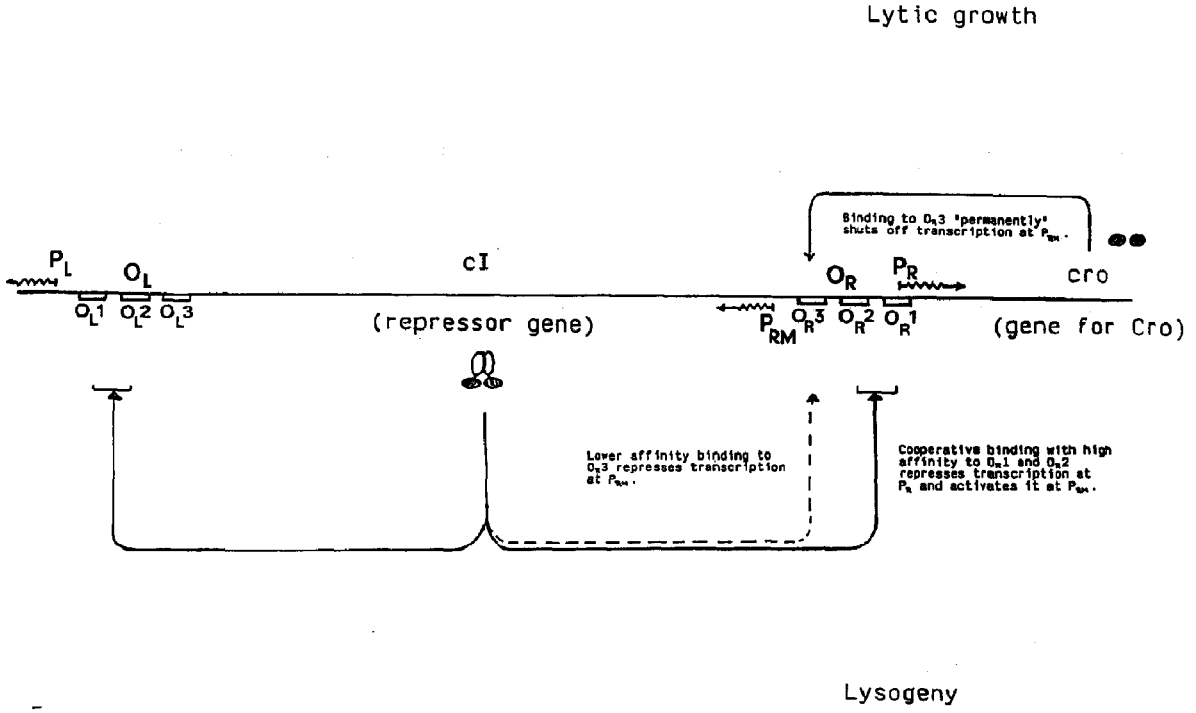

Fig. 1. Elements of the central regulatory interactions in temperate bacteriophages.

versible. At the thermodynamic level, the essence of this switch is summarized by the differential affinities in table 1 [2,3]. The measurements include sites in O_L, which regulates transcription of additional lytic functions. In structural terms, the key questions are: what features of the designs of two very similar proteins (the repressor and cro proteins of 434) permit them to bind with opposite rank orders of affinity to three very closely related DNA sequences? What aspects of the DNA sequence and structure are involved?

2. Crystallography

The repressor of bacteriophage 434 is a dimer of identical subunits, each of which folds into two domains [4]. The N-terminal domain mediates DNA binding. It comprises the first 69 residues of the polypeptide chain. We have studied this fragment ('R1–69'), prepared by proteolytic cleavage, and the 71-residue 434 Cro, both as free proteins and complexed with operator DNA. We have completed X-ray crystallographic analyses of R1–69 (to 2 Å resolution), of 434 Cro (to 2.5 Å), and of R1–69 bound to a 14 bp synthetic operator (to about 3.2 Å). Analysis of the structure of 434 Cro bound to the same synthetic operator (to about 4 Å) is in progress. The sequence of this synthetic operator ('14'-mer') is shown in table 1. We have also prepared crystals of R1–69 in complex with other synthetic operators, and work on these is in progress. This report focuses on our interpretation of the R1–69-14-mer complex [5]. Structural descriptions of free R1–69, of 434 Cro, and of the Cro-14-mer complex will be reported elsewhere.

3. R1–69 and 434 Cro

Fig. 2 shows a summary comparison of R1–69 and 434 Cro. The close similarity is evident. This similarity was anticipated from the high degree of

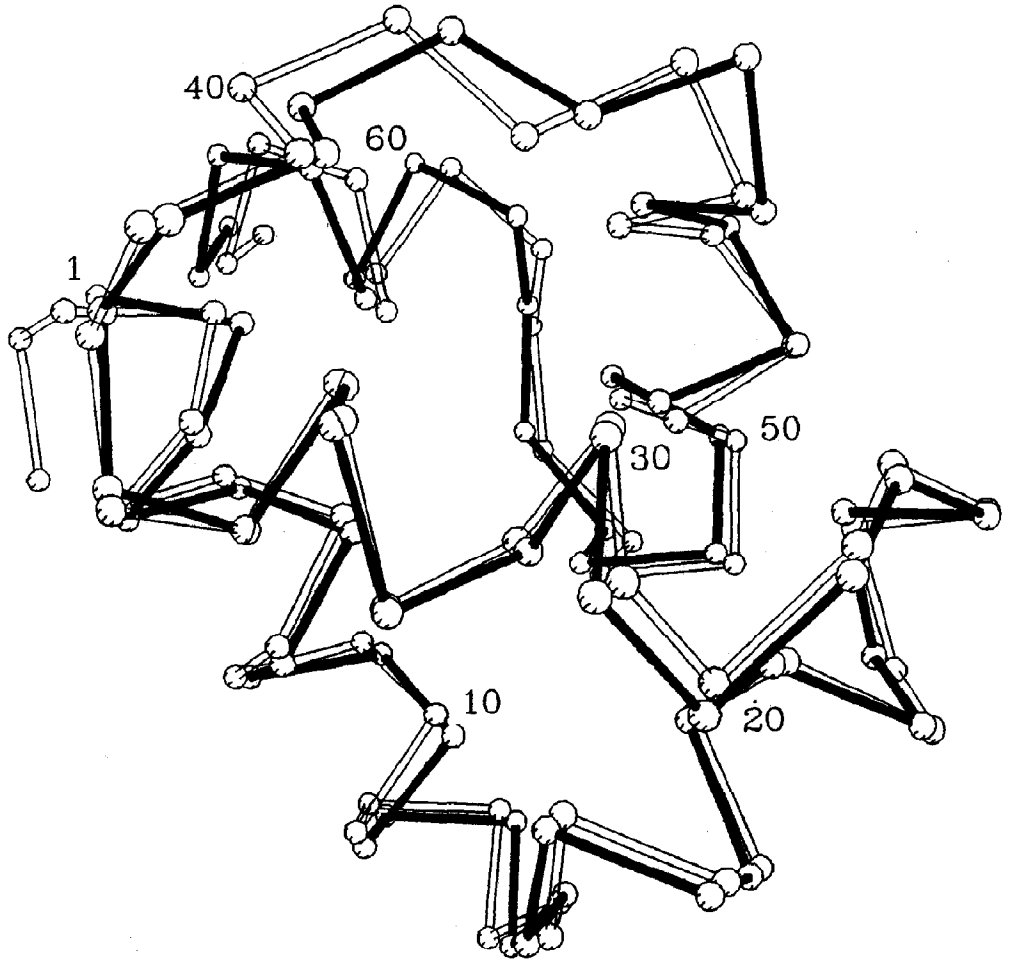

Fig. 2. Comparison of structures of R1–69 (solid bonds) and 434 Cro (open bonds). Only Cα positions are shown. Links are virtual bonds connecting these atoms. View is from point of view of DNA in a complex; α3 (residues 27–36) extends across the center of the diagram.

amino acid sequence identity. Both proteins contain the helix-turn-helix element first seen in the cro protein of phage λ [6] and in the *E. coli* CAP protein [7]. Unlike the λ repressor and Cro, the 434 proteins contain this element embedded in essentially identical domains. These domains are clusters of five α-helical segments. The part of the structure that forms helices 1–4 is similar to the corresponding part of the N-terminal domain of the λ repressor [8]. The extended N-terminal arm of the λ repressor, believed to embrace DNA in the complex, is missing in the 434 proteins, as is the long helix 5 [8].

4. The R1–69-14-mer complex

The complex of R1–69 and 14-mer is shown in fig. 3 [5]. Except for the reorientation of a few side chains that face DNA, the conformation of each R1–69 monomer is the same as that of free R1–69. The DNA in the complex is a B-type helix, and adjacent 14-mers stack so precisely on each other in the crystal that they mimic continuous B-DNA. One reason for this precise stacking is a contact between the N-terminus of α2 and P14 of the adjacent 14-mer; this contact is probably also found when R1–69 is bound to continuous DNA

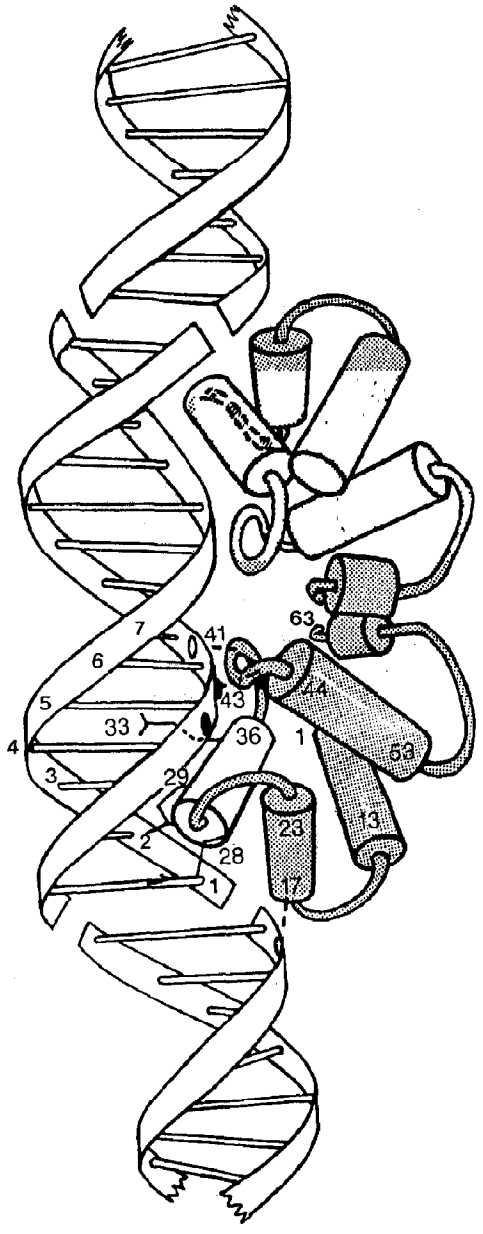

Fig. 3. Overview of the R1–69/14-mer complex. Helices in the protein are shown as cylinders; numbers indicate key residues as well as first and last residues of α1 – α4.

[9]. Another reason may be the favorable energetic contribution of base stacking. End-to-end stacking is found in many DNA crystals. The DNA helix is gently bent around the repressor domains. The B-helix is overwound in the center and somewhat underwound at the ends. A consequence of the combined bend and twist is a marked narrowing of the minor groove in the center. The mean twist is approx. 10.5 bp/turn. (It would be precisely 10.5, were the helix axis straight, since three successive 14-mers, related by a 3-fold screw axis, constitute a crystallographic repeat of 42 bp in four turns [4]. The slight right-handed writhe imparted by the curvature of the DNA leads to an average *local* twist of slightly less than 10.5 bp/turn.)

5. Interactions with DNA backbone

The configuration of the 14-mer appears to be fixed by a set of interactions between the protein and phosphate groups of the DNA backbone. In B-DNA, phosphates project significantly beyond the ribose rings, forming a series of protrusions that define the helical path. Four phosphates in each half-operator are in van der Waals contact with groups on the repressor domain. Three of these (P9, P10, and P14 of the adjacent 14-mer, here denoted 'P14-'; see fig. 4) fit precisely into shallow pockets on the protein surface, lined by one or more peptide -NH groups. These main-chain amides appear to form hydrogen bonds to phosphate oxygens. Adjacent arginine side chains also add positive charge to the vicinity of each of the sites. The site for P14-, which corresponds to the phosphate between the first and second base-pairs outside the operator in continuous DNA, is formed by the N-terminus of α2 and adjacent side chains; the sites for P9 and P10 are formed by the loop between α3 and α4. The fourth phosphate in contact (P11) lies against the turn between α2 and α3, but it remains more exposed than the other three. Ethylation of any one of these four phosphates interferes with repressor binding [9]. Interference from modification of P11 is less pronounced than P9, P10, or P14-, as expected from the extent of protein-phosphate interaction in each case.

6. Interactions with DNA bases

Docking of a repressor DNA-binding domain with an operator half-site, established by the fit to

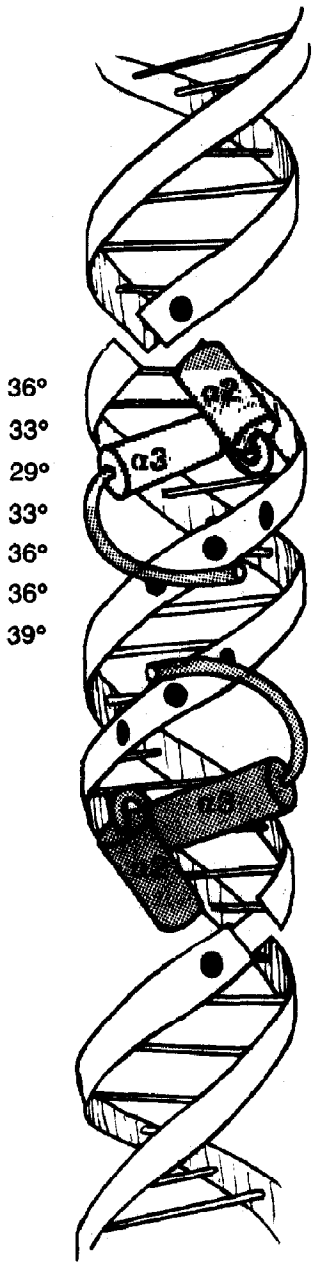

Fig. 4. Phosphate contacts and local DNA twist, indicated on a view normal to the view of fig. 3. In the protein, only α2 and α3 are shown.

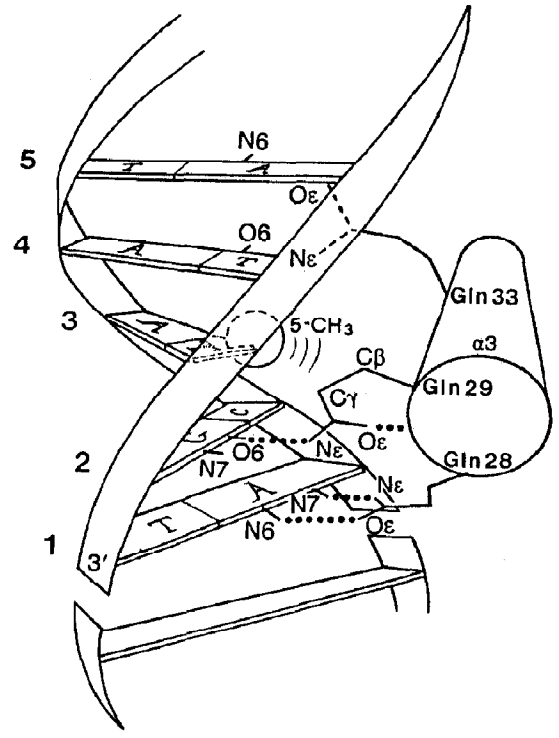

Fig. 5. Diagram of α3 and side-chain contacts in the major groove. Gln 28, Gln 29, and Gln 33 are shown. The van der Waals contact between Ser 30 and the methyl group of T11′ is not indicated. The nucleotide numbering scheme is: A1 and T14′ in base-pair 1, C2 and G13′ in base-pair 2, etc.

DNA backbone just described, positions α3 in the major groove. In principle, the side chains of residues 28–30, 32 and 33 could contact DNA bases. At the present resolution, the configurations of these side chains have been established, and assignment of van der Waals contacts and hydrogen bonds to DNA bases can be inferred from the geometry (fig. 5). Interactions with DNA bases can be made by three glutamine residues (28, 29 and 33), as well as by Ser 30. Glu 32 folds back away from the DNA to form hydrogen bonds with Gln 17. Gln 28 projects towards A1, so that it can form a bidentate hydrogen bond to the base. Gln 29 folds back toward the N-terminus of α3. Its Cβ and Cγ are in van der Waals contact with the methyl group of T12′ (base-pair 3); its Nε appears to donate hydrogen bonds to O6 and N7 of G13′ (base-pair 2). Glutamines 28 and 29 thus form a complementary surface for parts of base-pairs 1–3. No other sequences of bases can 'fit' this surface, and indeed these bases are conserved

in all naturally occurring 434 operators and in all synthetic operators with measurable affinity for 434 repressor [2,3]. Several further observations are consistent with this picture. (1) Mutation of Gln 28 to Ala produces a repressor that binds with reasonable affinity to an altered operator with T:A (instead of A:T) at base-pair 1 [10]. Model building suggests that the thymine methyl group can fit into a van der Waals pocket formed by the Cβ of Ala 28 and Cα of Gln 29. (2) If residues of the repressor of phage P22, homologous to that of 434, are substituted at positions 27–29 and 32, the hybrid protein binds selectively to P22 operators [11]. That is, P22-like residues of α3, docked onto DNA by the framework of a 434 repressor, are suitably complementary to P22 operator sequences. The amino acid residue corresponding to Gln 29 is valine. In a 434-like fit to DNA, we would expect it to make good van der Waals contact with a thymine methyl group in base-pair 3, but no hydrogen bonds (or other contacts) to base-pair 2. Indeed, the P22 base-pair corresponding to position 3 in 434 is also A:T, but the base-pair corresponding to position 2 is not conserved.

We do not yet fully understand the specificity at base-pairs 4 and 5. Gln 33 projects toward these base-pairs. It can make a (somewhat long) hydrogen bond with O4 of T11′ (base-pair 4), but no direct hydrogen bond to A10′ (base-pair 5 of the 14-mer in our crystals). Were base-pair 5 changed to G:C, a hydrogen bond to N4 of C10′ could also be made while maintaining the bond to T11′. Specificity at base-pair 4 is important for the crucial differential affinities of 434 repressor and Cro at O_R3. In O_R3, one half-site contains GC (instead of AT) at this position. The 434 repressor binds poorly to a 14-mer with GC substituted at position 4: the 434 Cro is indifferent to such a change. One principal difference between α3 of 434 repressor and 434 Cro is residue 33 (Gln in repressor, Leu in Cro). The properties of a mutant repressor with Ala at this position suggest that interaction with Gln 33 accounts for some but not all of the discrimination at base-pair 4 (Koudelka and Wharton, personal communication). A possible further source of specificity lies at the N-terminal end of α3. The Oγ of Ser 30 (hydrogen-bonded to Thr 27) is in van der Waals contact with the 5-methyl group of T11′ (base-pair 4). This contact could contribute to repressor specificity at base-pair 4, since any other base at 11′ would leave a 'hole' bounded by base 11′ and Ser 30 (Oγ). It could contribute to differential affinity of repressor and Cro only if the Cro/operator complex positions α3 somewhat differently. Elucidation of the structure of the Cro/14-mer complex should help in answering this question.

7. Indirect recognition of the central base-pairs

Base-pairs 6 and 7 face protein only through the narrow central minor groove [5]. The base composition at these positions does affect repressor binding [12]. The approximate rule is that operators with A:T or T:A at these positions have higher affinity for repressor than those with G:C or C:G. The only possible direct interaction with base-pairs 6 or 7 would be made by Arg 43, which projects into the minor groove in this region. The side chain is ill-defined in our map, but mutation of this residue to alanine does not alter the preference for A:T or T:A base-pairs (although it does uniformly lower affinity [12]. These observations suggest that the preference is due to the DNA configuration required when an operator is bound to a dimeric repressor. As described above, backbone contacts make the docking of each R1–69 domain quite precise. Moreover, contacts between the protein monomers involve interdigitated hydrophobic surfaces. Only an appropriate geometry at the center of the operator will permit optimal orientation of these protein dimer interactions while maintaining correct positioning of each monomer on DNA. A:T or T:A base-pairs are more conducive to the observed overwinding and narrowing of the minor groove than G:C or C:G base-pairs.

8. Conclusions

The 434 repressor-operator complex is a very simple structure. No protein arms envelop the DNA; no major bending wraps DNA around

protein. Molecular recognition in this complex has two aspects: (1) positioning of repressor on correctly configured DNA and (2) selection of the DNA sequence over which this positioning takes place. We observe that the first aspect is governed by the way in which protruding DNA phosphates fit into a set of shallow pockets on the repressor surface. These pockets are adjacent to positive charge, but the fit of a phosphate group appears to be determined particularly by hydrogen bonds to peptide-bond amide groups. Van der Waals contacts to ribose moieties may also affect positioning. We have shown that the second aspect of recognition is mediated both directly by the fit of $\alpha 3$ into the major groove and indirectly by the way in which base composition confers flexibility on the central base-pair region. The direct $\alpha 3$/base-pair interactions involve nonpolar van der Waals contacts as well as hydrogen bonds. Thymine methyl groups play a significant role. The indirect influence of central base-pair composition is a consequence of repressor binding as a dimer and of the inflexible docking of the repressor monomers on each half-site. The DNA configuration in the center of the operator must be consistent with the requirements of dimer contacts in the protein.

Acknowledgements

This work was supported by NIH grant GM-29109. J.E.A. was a Burroughs-Wellcome Fellow of the Life Sciences Research Foundation and G.B.K. and A.M. were Fellows of the Damon Runyon-Walter Winchell Cancer Fund.

References

1 M. Ptashne, A genetic switch (Cell Press, Cambridge, MA, 1986).
2 R.P. Wharton, E.L. Brown and M. Ptashne, Cell 38 (1984) 361.
3 R.P. Wharton, PhD Thesis, Harvard University (1985).
4 J.E. Anderson, M. Ptashne and S.C. Harrison, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 1307.
5 J.E. Anderson, M. Ptashne and S.C. Harrison, Nature 326 (1987) 846.
6 W.F. Anderson, D.H. Ohlendorf, Y. Takeda and B.W. Matthews, Nature 290 (1981) 754.
7 D.B. McKay and T.A. Steitz, Nature 290 (1981) 744.
8 C. Pabo and M. Lewis, Nature 298 (1982) 443.
9 F.D. Bushman, J.E. Anderson, S.C. Harrison and M. Ptashne, Nature 316 (1985) 651.
10 R.P. Wharton and M. Ptashne, Nature 326 (1987) 888.
11 R.P. Wharton and M. Ptashne, Nature 316 (1985) 601.
12 G.B. Koudelka, S.C. Harrison and M. Ptashne, Nature 326 (1987) 886.
13 A.D. Johnson, A.R. Poteete, G. Lauer, R.T. Sauer, G.K. Ackers and M. Ptashne, Nature 294 (1981) 217.