DNA-Binding Proteins

Y. Takeda, D. H. Ohlendorf
W. F. Anderson, B. W. Matthews*

Summary. The structures of three proteins that regulate gene expression have been determined recently and suggest how these proteins may bind to their specific recognition sites on the DNA. One protein (Cro) is a repressor of gene expression, the second (CAP) usually stimulates gene expression, and the third (lambda repressor) can act as either a repressor or an activator. The three proteins contain a substructure consisting of two consecutive alpha helices that is virtually identical in each case. Structural and amino acid sequence comparisons suggest that this bihelical fold occurs in a number of proteins that regulate gene expression, and is an intrinsic part of the DNA-protein recognition event. The modes of repression and activation by Cro and lambda repressor are understood reasonably well, but the mode of action of CAP is still unclear.

Recognition of DNA by proteins is not only of central importance in biology but also presents fascinating kinetic and structural problems. The chromosome of *Escherichia coli* contains about three million nucleotide base pairs (bp) and, as seen in the electron microscope, is essentially featureless. The site on the DNA to which lac repressor, for example, binds, consists of about 20 bp. Not only does lac repressor recognize this site and bind to it 10 million times more tightly than to the rest of the DNA, but it is estimated to reach its target site 1000 times more rapidly than anticipated by simple diffusion (*1, 2*). Although lac repressor has been studied extensively, its structure has not yielded to crystallographic analysis. However, the structures of three other sequence-specific DNA-binding proteins [Cro, CAP, and λ repressor] have now been determined and the result suggests how these proteins recognize DNA.

The principal characteristics of Cro (*3–6*), λ repressor (*7–11*), and CAP (*12–17*) are outlined in Table 1 (λ repressor is the product of the cI gene of bacteriophage λ and is alternatively referred to as cI). All three proteins are dimers, but have substantial differences in their overall structures. For both CAP and λ repressor, the respective polypeptides form two domains. In λ repressor, the amino-terminal domain binds to the DNA, whereas in CAP it is the carboxyl terminal part of the molecule that has this function.

Both Cro and λ repressor bind to six similar but nonidentical 17-bp sites on the chromosome of bacteriophage λ. The site at which Cro binds most tightly is the weakest binding site for λ repressor, and vice versa. The two proteins are involved in the adoption by the phage of either the lytic or the lysogenic mode of development (*18–22*). Although this selection process is complicated and not yet understood in its entirety, the respective roles of Cro and λ repressor have been analyzed in detail (*18–22*). Cro is a straightforward repressor of gene expression. It binds preferentially to the DNA at its highest affinity site (O_R3) and, when bound, prevents transcription from the adjacent repressor maintenance promoter, P_{RM}. The λ repressor is more versatile. In common with Cro it can act as a repressor but it can also stimulate the expression of its own gene. The protein CAP ("catabolite gene activator protein," also known as "cyclic AMP receptor protein") participates in the regulation of a number of genes in *E. coli* (*12–16*). In the presence of cyclic AMP (adenosine monophosphate), CAP promotes transcription of these genes (*12*), and, in some circumstances, can act as a negative regulator as well (*14*). The DNA sequence that CAP recognizes is approximately 15 nucleotides long and more than ten such sites are known in *E. coli* (*12, 14–16*).

Dr. Takeda is a professor in the Chemistry Department, University of Maryland, Baltimore County, Catonsville 21228. Dr. Ohlendorf is a research associate and Dr. Matthews is a professor of physics and director of the Institute of Molecular Biology, University of Oregon, Eugene 97403. Dr. Anderson is a professor in the Medical Research Council Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7.

Structures of DNA-Binding Proteins

A sketch of the structures of Cro (*6*), the amino-terminal domain of λ repressor (*11*), and the carboxyl-terminal domain of CAP (*17*) as determined from the respective crystal structures is shown in Fig. 1.

The structure of Cro is quite simple, consisting of three α helices (α_1, α_2, α_3) and a three-stranded antiparallel β sheet (*6, 23*). In the crystal, four polypeptides associate as a tetramer with approximate 222 symmetry. Ultracentrifugation indicates that Cro is dimeric in solution (*3*). This dimer is presumed to be the one shown in Fig. 1. Residues 55 to 61 of each monomer extend and lie against the surface of the other monomer. In particular, Phe^{58} makes intimate hydrophobic contact with its partner subunit. The carboxyl-terminal residues 62 to 66 are disordered in the crystals and, presumably, in solution as well.

The DNA-binding form of intact λ repressor is predominantly a dimer (Table 1). Proteolytic cleavage separates the protein into an amino-terminal domain of 92 residues and a carboxyl-terminal domain consisting of residues 132 to 236 (*7*). The amino-terminal domain can interact specifically with DNA and act as both a positive and negative regulator of transcription. Although monomeric in solution, the amino-terminal domain dimerizes as it binds to DNA and protects exactly the same bases against chemical modification as intact λ repressor (*8*). It is this amino-terminal domain whose structure has been determined by Pabo and Lewis (*11, 24*) and is shown in Fig. 1.

The structure contains five α helices with no β sheet. In the crystal, the two amino-terminal domains make contact via the fifth α helix in each subunit (Fig. 1). Studies of mutants of λ repressor suggest that a similar helix-helix contact may occur in intact λ repressor. In particular, there are mutations in this contact region that are known to interfere with the binding of intact repressor to DNA (*24, 25*).

The complex of CAP with cyclic AMP, that is, the DNA-binding form of CAP, was shown (*17, 26, 27*) to be a two-domain structure. Figure 1 includes only the carboxyl-terminal domains, that is, the presumed DNA-binding region. Not shown is the larger amino-terminal domain to which the cyclic AMP is bound

(*27*). The structure in the crystals is dimeric with the amino-terminal domains related by a local twofold axis. However, the carboxyl-terminal domains adopt somewhat different conformations and are not exactly twofold-related.

Models for DNA Binding

To date, no structure has been determined of a sequence-specific DNA-binding protein complexed with its target DNA sequence. Structural models for DNA-protein recognition therefore rely primarily on inspection of the uncomplexed proteins as seen in the respective crystal structures (Fig. 1).

For Cro protein, the 34-Å spacing between the twofold-related α_3 helices, together with their angle of tilt (Fig. 1), strongly suggests that these α helices bind within successive major grooves of right-handed Watson-Crick DNA as illustrated in Fig. 2 (*6, 23, 28*). It is presumed that the flexible carboxyl-terminal residues of Cro participate in DNA binding by lying along the minor groove.

This model for Cro binding is not only attractive stereochemically, it also is consistent with chemical protection and modification studies of the DNA (*21, 29*) as well as the protein (*23, 30*), and is supported by recent nuclear magnetic resonance (NMR) studies (*31*). A characteristic feature of the model is the match between the twofold symmetry of the protein and the (approximate) twofold sequence and spatial symmetry of the DNA binding site. It may also be noted that the concept that α helices of proteins bind within the major groove of DNA is not a recent one (*32*).

A similar mode of DNA binding has been proposed for λ repressor (*11, 24*). Here also there is a pair of twofold-related α_3 helices (Fig. 1) that are presumed to bind within successive major grooves of Watson-Crick DNA. Furthermore, the amino-terminal residues of λ repressor form two long "arms" with flexible ends that "wrap around" the DNA when the protein binds (*11, 33*). In this case the "arms" contact the major groove of the DNA. The presumed mode of DNA binding is consistent with DNA protection and modification studies (*21, 22, 24*). Also the model is well supported by the observed locations of mutations in λ repressor that alter DNA binding (Fig. 3) characterized by Sauer and co-workers (*25*).

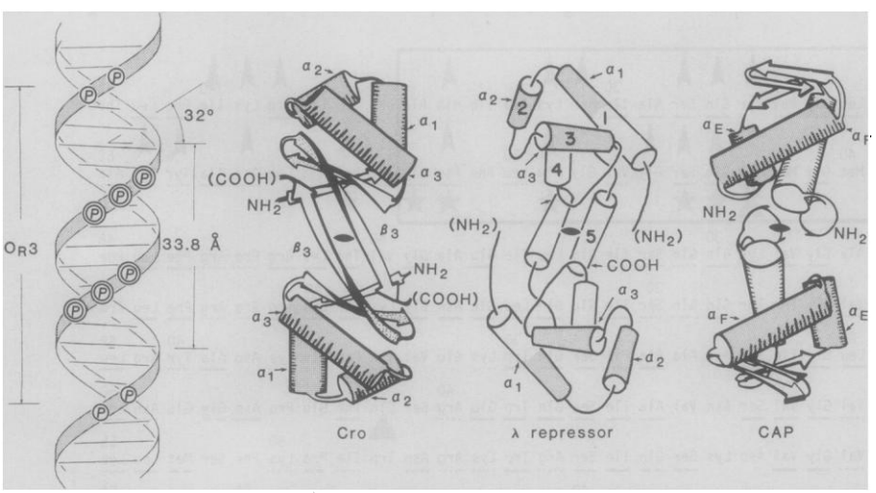
The mode of interaction of CAP with DNA is a controversial subject. As can be seen in Fig. 1, the α_F helices are 34 Å apart, but their tilt (compare Cro) is not obviously complementary to the grooves of right-handed B-form DNA. Faced with this dilemma, Steitz and co-workers (*17, 26, 27*) proposed that the DNA changes from right-handed to left-handed when CAP binds to its sequence-specific sites (*33a*). However, subsequent experiments designed to test this hypothesis indicate that there is no unwinding of the DNA when CAP binds to such sites (*34*). Pabo and Lewis (*11*) have suggested that CAP could bind to right-handed DNA. Allowing for some flexibility of the protein, together with bending of the DNA as may occur for Cro (*28*), it appears that a satisfactory complex of CAP with right-handed DNA can be built (*35*).

Sequence similarities in DNA-binding proteins. As a result of the structure determinations of Cro, CAP, and λ repressor, it has become apparent that these three proteins have features in common that presumably extend to a

Table 1. Double-strand DNA-binding proteins.

Protein	Action	Monomer	Active form	Domain structure*
λ-Cro (Cro)	Repressor	66 amino acids	Dimer	
λ-Repressor	Repressor or activator	236 amino acids	Dimer	(N) (C)
Catabolite gene activator (CAP or CRP)	Activator or repressor	209 amino acids	Dimer	(N) (C)

*DNA is represented by the black indicators.



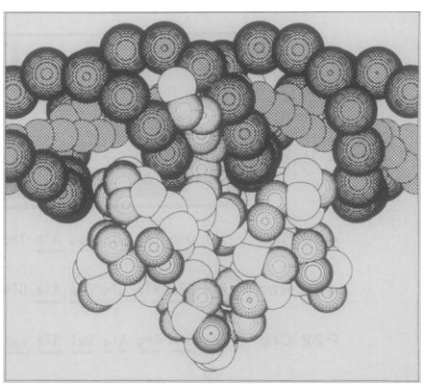
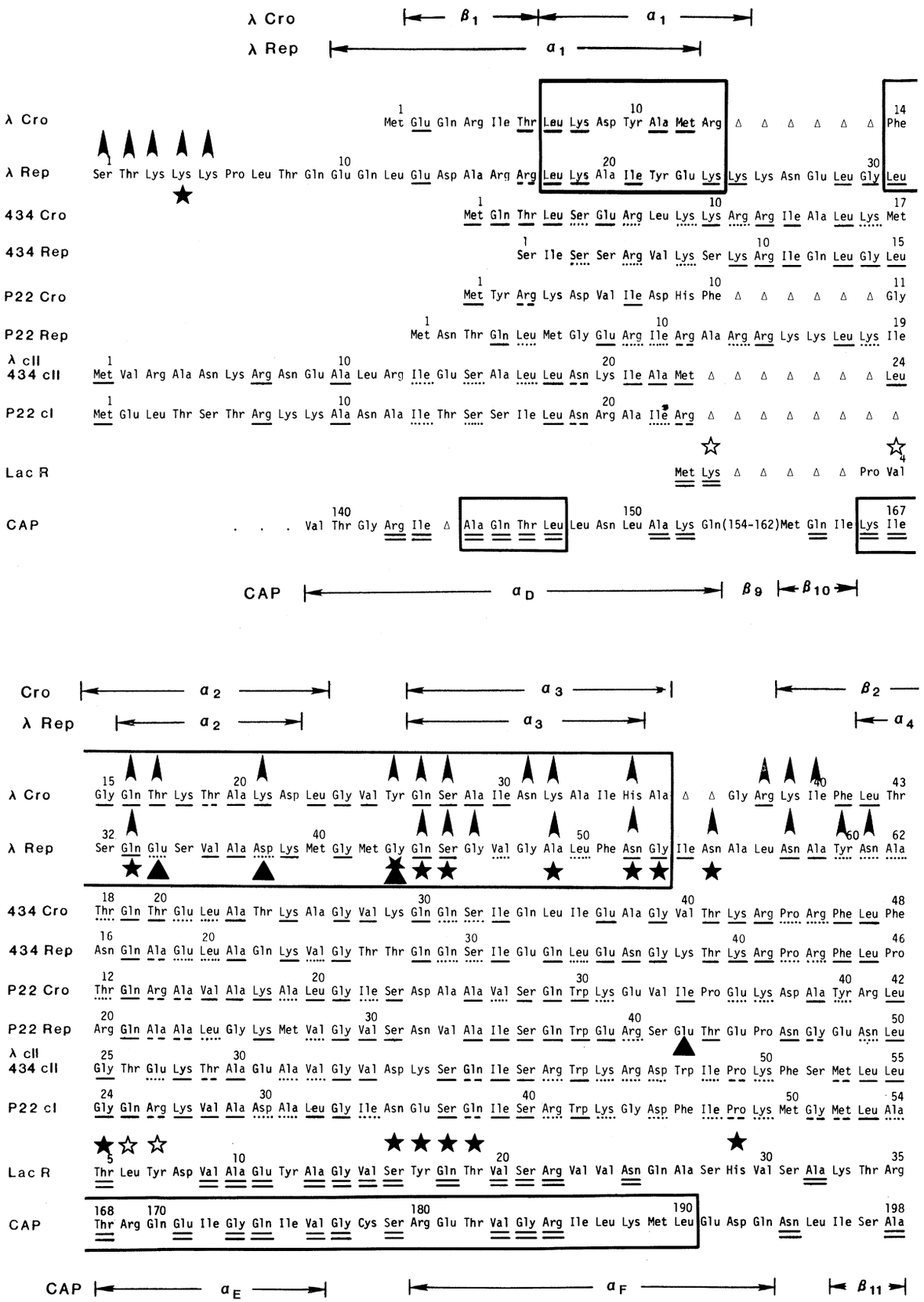


Fig. 1 (left). Schematic drawing of a segment of Watson-Crick B form DNA together with dimers of Cro, λ repressor amino-terminal domains, and CAP carboxyl-terminal domains viewed down their respective twofold symmetry axes. The parts of the α_1, α_2, and α_3 (or α_D, α_E, α_F) helices that spatially correspond are shaded. DNA phosphates whose ethylation interferes with binding of both λ repressor and Cro are indicated by the letter P within a double circle. Phosphates whose ethylation effects λ repressor (and also P22 repressor) binding, but not Cro, are indicated by a P in a single circle [based on (*6, 11, 17, 21, 29, 41, 45*]. Fig. 2 (right). Stylized drawing showing the complementarity between the structure of Cro repressor protein and DNA. In the presumed sequence-specific complex, the protein is assumed to move closer to the DNA, with the α_3 helices penetrating further into the major grooves of the DNA than is shown in the figure. The carboxyl-terminal residues of the protein are presumed to bind in the vicinity of the minor groove of the DNA. In order to maximize the contacts between Cro and DNA, the protein may undergo a hinge-bending motion or the DNA may bend (as shown), although these are not essential features of the model. The DNA is represented stylistically by large dotted spheres centered at the phosphate positions and small dotted spheres that follow the bottom of the major groove. In the protein, each residue is represented by a single sphere. Acidic residues have solid concentric circle shading, basic residues have broken circle shading, uncharged hydrophilic residues have dotted circle shading, and hydrophobic residues have no shading [after (*6, 28*)].

number of other DNA-binding proteins.

The suggestion that several DNA-binding proteins might have structural similarities came first from comparisons of their amino acid sequence (*36–39*). In some cases, such as Cro and λ repressor, the sequence homology is poor, and was not apparent on first inspection (*4, 9*). However, with additional sequences available, the overall homology becomes obvious (Fig. 3). The sequence homology includes not only repressors and activator proteins from different phages, but also other DNA-binding proteins such as the lac repressor from *E. coli*.

The region of best sequence homology occurs within the parts of the sequences that align with the α_2 and α_3 helices of



Cro, that is, within the part of the protein that has been proposed to interact with the DNA. Thus, it was reasonable to infer that the homologous proteins contained an α-helical DNA-binding supersecondary structure similar to the α_2–α_3 fold seen in Cro (*36, 37*). An independent series of sequence comparisons, made in the light of the known structure of λ repressor, led to a similar conclusion (*38*). The locations of known mutants of lac repressor are consistent with such a hypothesis (Fig. 3) (*37*), and additional support in this case has subsequently come from NMR studies (*40*).

Structural similarities in DNA-binding proteins. Apart from the above relationships, Cro and CAP have a striking structural correspondence in their presumed DNA-binding regions (*41*). The three α helices (α_D, α_E, α_F) in the carboxyl-terminal domain of CAP can be approximately superimposed on the α_1, α_2, and α_3 helices of Cro (Figs. 1 and 3). For the α_E–α_F and α_2–α_3 helical units the superposition is striking. There are 24 α carbons in the respective units that superimpose with an average discrepancy of 1.1 Å. An exhaustive search through all protein structures in the Brookhaven Data Bank failed to find a similar two-helical unit (Fig. 4). A subsequent, extended search for a pair of helices superimposable on the α_2 and α_3 helices of Cro, allowing the linkage between the two helices to be of arbitrary length, also failed to find a similar structure (*42*). After the gene sequence was determined and amino acid sequence of CAP (*13*) was inferred, it was shown that there is an amino acid sequence homology between CAP and other DNA-binding proteins corresponding with the observed structural homology between CAP and Cro (*43, 44*).

For Cro and λ repressor, their α_2 and α_3 helices, and parts of their α_1 helices as well, spatially superimpose (*45*) (Figs. 1 and 3). Again, as with Cro and CAP, it is the α_2–α_3 helical units of the two proteins that have virtually identical conformations (Fig. 4).

The alignment of the sequences of the DNA-binding proteins included in Fig. 3 is based in part on sequence homology and also on structural homology between Cro, CAP, and λ repressor. The residues that are buried tend to be most strongly conserved. Amino acid sequence comparisons made on the basis of the known structures of Cro, CAP, and λ repressor suggest that the two-helical unit may occur in additional DNA-binding proteins (not included in Fig. 3) that are involved in gene regulation at the level of transcription (*23, 24, 38, 43, 44*).

Fig. 3. Alignment of a series of DNA-binding proteins based on sequence homology (*36–38*) and correspondence of the three-dimensional structures of Cro, CAP, and λ repressor (*41, 45*). The first eight sequences are DNA-binding proteins from bacteriophages. Within this set, residues identical in two or more sequences are underlined. Residues in the final two sequences (lac repressor and CAP) that agree with the phage proteins have a double underline. Residues that have similar three-dimensional structures in Cro and λ repressor (*45*) or in Cro and CAP (*41*) are boxed. Residues in Cro and λ repressor that are presumed to contact DNA are capped by arrowheads. Additional presumed contacts occur for residues near the carboxyl terminus of Cro. The solid stars underneath the λ repressor sequence show locations where amino acid substitutions interfere with the function of the protein (*25*). (Additional substitutions occur at Leu^{65}, Phe^{76}, Ser^{77}, and Ile^{84}. Substitutions of buried residues and replacements by proline are not included.) Similarly, the locations of residues of lac repressor for which all known mutants dramatically reduce DNA binding, but do not simply destabilize the protein, are indicated by solid stars. The open stars indicate locations where amino acid substitutions may reduce DNA binding, or where the reduction in DNA-binding affinity is weak (*37*). Locations in λ repressor and P22 repressor at which amino acid substitutions result in nonstimulation of transcription are indicated by triangles (*52, 53*). Approximate locations of α helices and β sheets in the three known structures are shown. (The definition of the ends of an α helix or β sheet is somewhat arbitrary, especially for an unrefined structure. Therefore slight discrepancies in the extent of α helices, as for the α_2 and α_3 helices of Cro and λ repressor, do not necessarily imply different conformations or hydrogen bonding patterns at the ends of these helices.) References to the amino acid and gene sequences are given in (*37, 38, 44*). In a number of cases the amino acid sequence is inferred from the gene sequence and the amino terminal residue of the protein is uncertain. The abbreviations for the amino acid residues are: Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

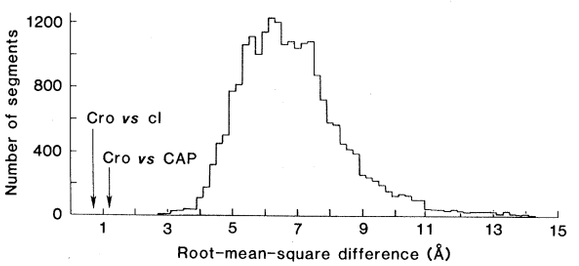


Fig. 4. Histogram showing the result of a search through all protein structures in the Brookhaven Data Bank for a bihelical unit similar to the α_2–α_3 unit as initially seen in Cro and CAP (*41*). $R_{C\alpha}$ is the root-mean-square distance between the 24 α-carbon atoms in the α_2–α_3 unit of Cro and successive 24 α-carbon segments from all known structures, aligned so as to minimize $R_{C\alpha}$. Values of $R_{C\alpha}$ for the α_2–α_3 segment of Cro compared with the α_E–α_F segments of CAP (*41*) and λ repressor (*45*) are shown.

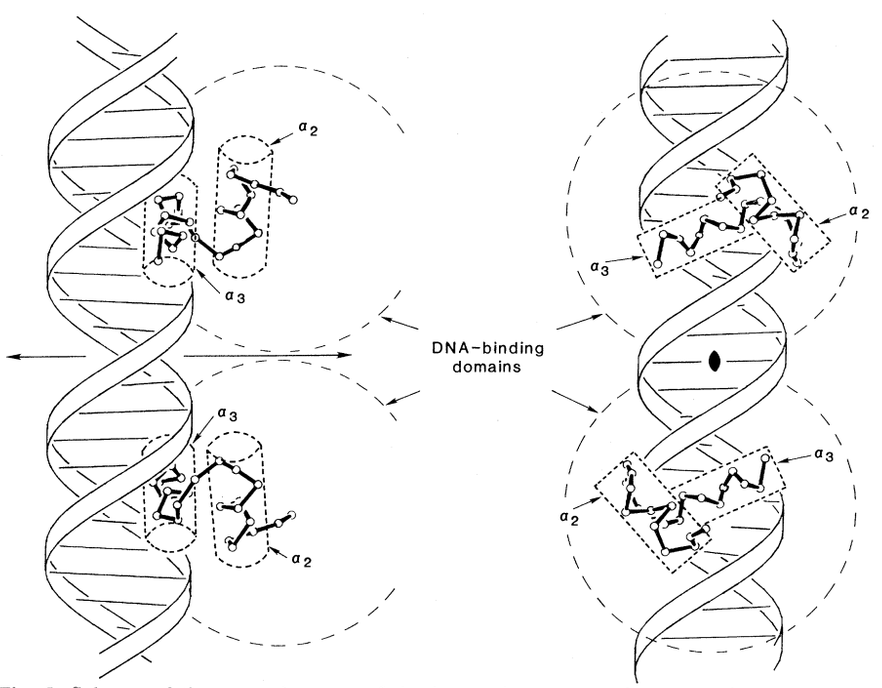


Fig. 5. Scheme of the general nature of the interaction presumed to occur in many DNA-regulatory proteins between a common α_2–α_3 helical unit and right-handed B-form DNA. At left is a "side view" with the twofold axis of symmetry (arrowed) extending from left to right. On the right the view is "face on" (*44*).

Similarities in DNA binding. The amino acid sequence comparisons and the structural comparisons both point to a special role for the two-helical "α_2–α_3" unit in DNA recognition and binding. The mode of interaction of this unit with DNA, as inferred from the structure of Cro (*6*), is sketched in Fig. 5. The presumed mode of binding of the α_2–α_3 helical unit in λ repressor is very similar although not identical (*11, 24, 45*). The α_3 helix occupies the major groove of the DNA with its amino acid side chains positioned so as to make sequence-specific interactions with the exposed parts of the DNA base pairs. Side chains of the α_2 helix are also presumed to contact the DNA, these interactions being primarily to the phosphate backbone.

It is reasonable to anticipate that similar modes of DNA binding will be found for a number of other gene regulatory proteins whose sequences have been shown to be homologous with those included in Fig. 3. Chemical and enzymatic (deoxyribonuclease footprinting) protection experiments have indicated the sizes of the recognition sites for such proteins to be about two turns of the DNA helix (usually 15 to 20 nucleotides). This is the DNA length which would be expected to be covered by the DNA recognition super-secondary structure.

The structural superposition of Cro on λ repressor (*45*) shows that the short loop connecting the α_1 and α_2 helices of λ repressor (Lys^{24} to Gly^{30}) is absent in Cro (Fig. 3). In the model for λ repressor bound to DNA (*11, 24*), this loop is located near the outer part of the binding site and might account for the different phosphate ethylation pattern for λ repressor and Cro. The phosphates whose ethylation interferes with Cro binding are located near the center of the operator site, whereas λ and P22 repressors have four additional phosphate contacts in the outermost parts of the binding site (Fig. 1). These additional contact sites, not seen for λ Cro, might be due to the "α_1–α_2 loop" present in the two repressors but absent from λ Cro (Fig. 3) (*45*).

Although Cro and λ repressor bind to the same sites on the DNA (albeit with different relative affinities), the amino acid sequences of their α_3 "recognition helices" are substantially different (Fig. 3). In contrast, comparison of the three-dimensional structures of the two proteins showed that the backbone conformations of their respective α_2–α_3 units were practically identical (*45*). However, the structural comparison also showed that there are stereochemical restrictions that prevent these units from binding to the DNA in exactly the same way. Cro and λ repressor apparently use a non-identical set of amino acid side chains organized in a somewhat different spatial arrangement to recognize the same DNA sequence. This suggests that there is not a simple one-on-one recognition code between amino acids and bases (*24, 36, 45*).

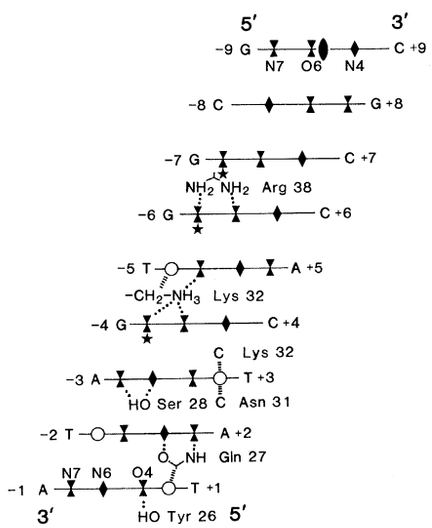


Fig. 6. Schematic representation of the presumed sequence-specific interactions between Cro and the parts of the base pairs exposed within the major groove of the DNA. The direction of view is imagined to be directly into the major groove of the DNA with the base pairs seen edge-on. The dyad symbol within the topmost base pair indicates the center of the overall 17-base-pair binding region. The symbols are as follows: ⧗, hydrogen bond acceptor: ◆, hydrogen bond donor; ○, methyl group of thymine; ★, guanine N-7 which is protected from methylation when Cro is bound (*29*). Presumed hydrogen bonds between Cro side chains and the bases are indicated (····). Apparent van der Waals contacts between Cro and the thymine methyl groups are shown (IIII). [From (*28*), courtesy of *Nature (London)*]

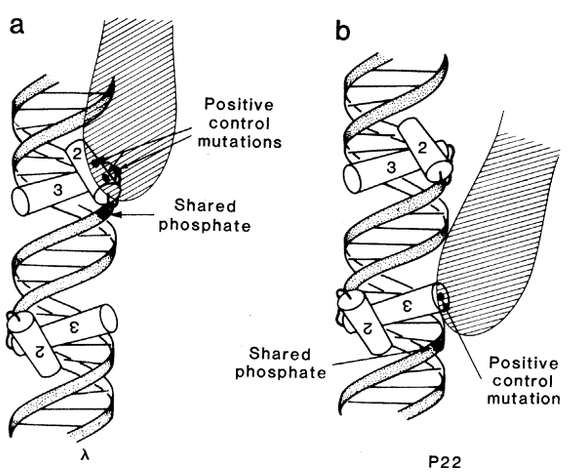


Fig. 7. (a) Schematic diagram of the relation between the α_2–α_3 helical unit of λ repressor and polymerase. (b) Similar diagram showing the relation between the presumed α_2–α_3 helical unit of P22 repressor and polymerase. In each case, polymerase is drawn shaded and the locations of the positive control mutations are shown. The "shared" phosphate is believed to be contacted by both bound repressor and bound polymerase. There is a similarity between the presumed alignment on the DNA of the α_2–α_3 helices of λ repressor (this figure) and that assumed for Cro (Fig. 5). [Courtesy of *Cell* (*53*)]

Recognition of Specific Sites

The above models for DNA-protein interaction are based in large part on an overall structural complementarity between the protein and the DNA. An underlying assumption of such models is that the DNA does not change its conformation very much when the protein is bound. This has been experimentally verified for lac repressor (*46*), λ repressor (*47*), and CAP (*34*). The complementarity between the shape of the protein and the shape of the DNA is also consistent with the observation that these proteins bind to nonspecific sites on the DNA ($K_D \sim 10^{-7}$ to $10^{-8} M$) in addition to their specific sites ($K_D \sim 10^{-10}$ to $10^{-13} M$).

In order to obtain an insight into sequence-specific recognition, model building and energy refinement were used to develop a detailed model for the complexes between Cro and DNA (*28*). This model is consistent with the known affinities of Cro for its six binding sites on λ DNA and for mutant sites as well. The model suggests that the recognition of a specific base sequence on the DNA is due, in large part, to a complementary network of hydrogen bonds between amino acid side chains of the protein and DNA base-pair atoms exposed within the grooves of the DNA (*48–50*).

The hydrogen bond network that is presumed to exist between Cro and its tightest known binding site, O_R3, is shown schematically in Fig. 6 (*28*). In the figure, the successive base pairs are imagined to be seen edge-on, with all the possible hydrogen bond acceptor and donor atoms indicated. Atoms that do not hydrogen bond to the protein are presumed to remain hydrogen bonded to

solvent. One striking feature of the model is the multiple hydrogen bonding of the amino acid side chains, for example, Arg^{38}, Lys^{32}, Gln^{27}, and even Ser^{28} (Fig. 6). Such bi- and multidentate interactions provide a clear rationale for enhancing the specificity of DNA-protein recognition (*48, 49*). Also, as has been shown for lac repressor, hydrophobic interactions can play an important part in recognition (*50*).

While protein-DNA interactions of the sort shown in Fig. 6 are presumed to be responsible for recognition of a specific base sequence, it is understood that the overall energy of interaction of the complex comes primarily from interactions with the DNA that do not depend on the base sequence (*49*). In the Cro-DNA model there are many such contacts, including about ten potential ionic interactions with the phosphate backbone (*28*). Model building suggests that when Cro is moved 4 Å or so away from the DNA, most of the hydrogen bonds that are involved in sequence-specific recognition would be broken, but many of the ionic interactions could be retained (Fig. 2). Such ionic interactions might be utilized by the protein in sliding along the DNA to reach the target site (*28, 51*).

Repression and Activation

The structural studies of Cro and λ repressor are consistent with the established idea that these proteins repress gene expression by binding at a site on the DNA that sterically prevents access of RNA polymerase to initiation sites for transcription.

Structural, genetic, and kinetic analyses of λ repressor in the laboratories of Ptashne and McClure indicate that this protein activates (or "positively controls") gene expression by a direct favorable contact with RNA polymerase (*21, 22, 52–54*). Three mutants of λ repressor that are deficient in positive control but do not greatly interfere with DNA binding, per se, have been isolated. The mutant proteins have amino acid alterations at positions 34(Glu → Lys), 38(Asp → Asn), or 43(Gly → Arg) (a decrease of negative charge in each instance) (*52, 53*). In terms of the three-dimensional structure of λ repressor these amino acids are in the α_2 helix or in the connection between the α_2 and α_3 helices (Figs. 1 and 3). In the proposed model for λ repressor bound to DNA the residues are clustered together on the surface that had been inferred from DNA protection studies to be a close contact between λ repressor and polymerase (Fig. 7) (*11, 22, 24, 52, 53*). Similar protection experiments for phage P22 had indicated that in this phage the distance from the relevant P22 repressor binding site to the P22 promoter (P_{RM}) is shorter than the corresponding distance in bacteriophage λ (*22*). Now, a mutant of P22 repressor defective in positive control has been isolated, and the substitution identified occurs at position 42 (Glu → Lys) (*53*). If it is assumed, on the basis of the sequence homology, that the structure of P22 repressor resembles the known structures of λ repressor and λ Cro, then Glu 42 would be located at the carboxyl end of the α_3 helix (Figs. 3 and 7b). This is not close to the presumed RNA polymerase contact region seen in λ repressor (Fig. 7a), in fact it is on the opposite side of the molecule. However, in a dimer of P22 repressor, residue 42 of the twofold-related monomer is exactly placed to contact polymerase in such a way that the distance from the repressor to P_{RM} is explained (Fig. 7b) (*22, 53*).

The evidence is therefore compelling that λ repressor and P22 repressors have acidic residues that are located in different parts of the respective structures and apparently make contact with the same "positive patch" on polymerase. This mechanism for the activation of transcription could well apply to other positive regulators. In the case of CAP stimulation of the gal, araC, and cat operons, the CAP binding site is in the "−35" region, that is, about 35 bases from the initiation of transcription (*16*), a distance comparable with P22 repressor. Also CAP has acidic residues (Glu^{191} and Asp^{192}) at positions that correspond to Glu^{42} of P22 repressor (Fig. 3), and might therefore make contact with polymerase. On the other hand, the distance from the CAP binding site to the origin of transcription varies in different operons. In particular, CAP binds to the −62 region in lac and the −85 to −90 region in araBAD and deo operons. Therefore it remains to be determined whether CAP stimulates transcription by making direct contact with polymerase, or if activation occurs by some other mechanism (*12, 14*).

References and Notes

1. J. H. Miller and W. S. Reznikoff, Eds., *The Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1978).
2. A. D. Riggs, S. Bourgeois, N. Cohn, *J. Mol. Biol.* **53**, 401 (1970); O. G. Berg, R. B. Winter, P. H. von Hippel, *Trends Biochem. Sci.* **7**, 52 (1982).
3. Y. Takeda, A. Folkmanis, H. Echols, *J. Biol. Chem.* **252**, 6177 (1977); Y. Takeda, *J. Mol. Biol.* **127**, 177 (1979).
4. N. W. Hsiang, R. D. Cole, Y. Takeda, H. Echols, *Nature (London)* **270**, 275 (1977).
5. T. M. Roberts, H. Shimatake, C. Brady, M. Rosenberg, *ibid.*, p. 274.
6. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *ibid.* **290**, 654 (1981).
7. C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1608 (1979).
8. R. T. Sauer, C. O. Pabo, B. J. Meyer, M. Ptashne, K. C. Backman, *Nature (London)* **279**, 396 (1979).
9. R. T. Sauer and R. Anderegg, *Biochemistry* **17**, 1092 (1978).
10. R. T. Sauer, *Nature (London)* **276**, 301 (1978).
11. C. O. Pabo and M. Lewis, *ibid.* **298**, 443 (1982).
12. G. Zubay, D. Schwartz, J. Beckwith, *Proc. Natl. Acad. Sci. U.S.A.* **66**, 104 (1970); M. Emmer, B. DeCrombrugghe, I. Pastan, R. Perlman, *ibid.*, p. 480; W. Epstein, L. B. Rothman-Denes, J. Hesse, *ibid.* **72**, 2300 (1975); J. Houmard and G. R. Drapeau, *ibid.* **69**, 3506 (1972); W. Gilbert, in *RNA Polymerase*, R. Losick and M. Chamberlin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1976), p. 193; W. S. Reznikoff and J. N. Abelson, in (*1*), p. 221; R. H. Ebright, in *Molecular Structure and Biological Function*, J. Griffen and W. Duax, Eds. (Elsevier, Amsterdam, 1982), p. 91; F. R. Salemme, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5263 (1982).
13. H. Aiba, S. Fujimoto, N. Ozaki, *Nucleic Acids Res.* **10**, 1345 (1982); P. Cossart and B. Gicquel-Sanzey, *ibid.*, p. 1363.
14. S. Adhya and S. Gorges, *Cell* **29**, 287 (1982); H. Aiba has observed that CAP negatively regulates its own synthesis [*ibid.* **32**, 141 (1983)]. Thus, all three proteins, Cro, λ repressor, and CAP are autorepressors.
15. CAP protects about 25 nucleotides against deoxyribonuclease digestion, but the recognition sequence is about 15 nucleotides in length.
16. S. F. L. LeGrice and H. Matzura, *J. Mol. Biol.* **150**, 185 (1981); S. Ogden, D. Haggerty, C. M. Stoner, D. Kol, R. Scheif, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3346 (1980); T. Taniguchi, M. O'Neill, B. deCrombrugghe, *ibid.* **76**, 590 (1979); A. Schmitz, *Nucleic Acids Res.* **9**, 277 (1981).
17. D. B. McKay and T. A. Steitz, *Nature (London)* **290**, 744 (1981).
18. M. Ptashne, K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. J. Meyer, R. T. Sauer, *Science* **194**, 156 (1976).
19. H. Echols, in *The Molecular Genetics of Development*, W. Loomis and T. Leighton, Eds. (Academic Press, New York, 1980), p. 1.
20. I. Herskowitz and D. A. Hagen, *Annu. Rev. Genet.* **14**, 399 (1980).
21. M. Ptashne *et al.*, *Cell* **19**, 1 (1980).
22. A. D. Johnson, A. R. Poteete, G. Lauer, R. T. Sauer, G. K. Ackers, M. Ptashne, *Nature (London)* **294**, 217 (1981).
23. B. W. Matthews, D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 427 (1982); B. W. Matthews, D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, *Trends Biochem. Sci.* **8**, 25 (1983).
24. M. Lewis, A. Jeffrey, R. Ladner, M. Ptashne, J. Wang, C. O. Pabo, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 435 (1982).
25. H. C. M. Nelson, M. H. Hecht, R. T. Sauer, *ibid.*, p. 441; M. H. Hecht, H. C. M. Nelson, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2676 (1983).
26. T. A. Steitz, I. T. Weber, J. B. Matthew, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 419 (1982).
27. D. B. McKay, I. T. Weber, T. A. Steitz, *J. Biol. Chem.* **257**, 9518 (1982).
28. D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, B. W. Matthews, *Nature (London)* **298**, 718 (1982).
29. A. D. Johnson, B. J. Meyer, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1783 (1978).
30. Y. Takeda, unpublished data.
31. K. T. Arndt, F. Boschelli, J. Cook, Y. Takeda, E. Tecza, P. Lu, *J. Biol. Chem.* **258**, 4177 (1983).
32. G. Zubay and P. Doty, *J. Mol. Biol.* **1**, 1 (1959); K. Adler *et al.*, *Nature (London)* **237**, 322 (1972).
33. C. O. Pabo, W. Krovatin, A. Jeffrey, R. T. Sauer, *Nature (London)* **298**, 441 (1982).

33a. *Note added in proof:* T. A. Steitz has informed us (personal communication) that he now favors a model with CAP bound to right-handed rather than left-handed DNA.

34. A. Kolb and H. Buc, *Nucleic Acids Res.* **10**, 473 (1982).
35. D. H. Ohlendorf and B. W. Matthews, *Annu. Rev. Biophys. Bioeng.* **12**, 259 (1983).
36. W. F. Anderson, Y. Takeda, D. H. Ohlendorf, B. W. Matthews, *J. Mol. Biol.* **159**, 745 (1982); R. Grosschedl and E. Schwarz, *Nucleic Acids Res.* **6**, 867 (1979).
37. B. W. Matthews, D. H. Ohlendorf, W. F. Anderson, Y. Takeda, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1428 (1982).

38. R. T. Sauer, R. R. Yocum, R. F. Doolittle, M. Lewis, C. O. Pabo, *Nature (London)* **298**, 447 (1982).
39. B. von Wilken-Bergmann and B. Müller-Hill, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2427 (1982).
40. H. Nick *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 218 (1982); H. Nick *et al.*, *J. Mol. Biol.* **161**, 417 (1982); K. Arndt, H. Nick, F. Boschelli, P. Lu, J. Sadler, *ibid.*, p. 439; R. Kaptein and K. Wuthrich, personal communication.
41. T. A. Steitz, D. H. Ohlendorf, D. B. McKay, W. F. Anderson, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3097 (1982).
42. D. H. Ohlendorf and B. W. Matthews, unpublished data.
43. I. T. Weber, D. B. McKay, T. A. Steitz, *Nucleic Acids Res.* **10**, 5085 (1982).
44. D. H. Ohlendorf, W. F. Anderson, B. W. Matthews, *J. Mol. Evol.* **19**, 109 (1983).
45. D. H. Ohlendorf, W. F. Anderson, M. Lewis, C. O. Pabo, B. W. Matthews, *J. Mol. Biol.*, in press.
46. J. C. Wang, M. D. Barkley, S. Bourgeois, *Nature (London)* **251**, 247 (1974).
47. T. Maniatis and M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1531 (1973).
48. N. C. Seeman, J. M. Rosenberg, A. Rich, *ibid.* **73**, 804 (1979); C. Hélene, *FEBS Lett.* **74**, 10 (1977).
49. P. H. von Hippel, in *Biological Regulation and Development*, R. F. Goldberger, Ed. (Plenum, New York, 1979), p. 279.
50. M. H. Caruthers, *Acc. Chem. Res.* **13**, 155 (1980).
51. O. G. Berg, R. B. Winter, P. H. von Hippel, *Biochemistry* **20**, 6929 (1981); R. B. Winter and P. H. von Hippel, *ibid.*, p. 6948. R. B. Winter, O. G. Berg, P. H. von Hippel, *ibid.*, p. 6961.
52. L. Guarente, J. S. Nye, A. Hochschild, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 274 (1982).
53. A. Hochschild, N. Irwin, M. Ptashne, *Cell* **32**, 319 (1983).
54. D. K. Hawley and W. R. McClure, *J. Mol. Biol.* **157**, 493 (1982); *Cell* **32**, 327 (1983).
55. We thank R. T. Sauer for communicating results in advance of publication and C. O. Pabo and M. Ptashne for helpful comments on the manuscript. Supported in part by NIH grants GM28138 and GM30894 (Y.T.) and GM20066 (B.W.M.), a grant from the M. J. Murdock Charitable Trust, NSF grant PCM-8014311 (B.W.M.), a grant from the Canadian Medical Research Council through the MRC Group on Protein Structure and Function (W.F.A.), and an NIH postdoctoral fellowship (D.H.O.).
* Address correspondence to B. W. Matthews.

Science, Risk, and Public Policy

William D. Ruckelshaus

We are now in a troubled and emotional period for pollution control; many communities are gripped by something approaching panic and the public discussion is dominated by personalities rather than substance. It is not important to assign blame for this. I appreciate that people are worried about public health and about economic survival, and legitimately so, but we must all reject the emotionalism that surrounds the current discourse and rescue ourselves from the paralysis of honest public policy that it breeds.

Summary. A climate of fear now dominates the discussion of environmental issues. The scientific community can help alleviate this fear by making a greater effort to explain to the public the uncertainties involved in estimates of risk. Current statutory mandates designed to protect public health both demand levels of protection that technology cannot achieve and are uncoordinated across government agencies. A common statutory framework for dealing with environmental risks is needed. In addition, care must be taken to separate the scientific process of assessing risk from the use of such assessments, together with economic and policy considerations, in the management of risks through regulatory action.

I believe that part of the solution to our distress lies with the idea that disciplined minds can grapple with ignorance and sometimes win: the idea of science. We will not recover our equilibrium without a concerted effort to more effectively engage the scientific community. Frankly, we are not going to be able to emerge from our current troubles without a much improved level of public confidence. The polls show that scientists have more credibility than lawyers or businessmen or politicians, and I am all three of those. I need the help of scientists.

This is not a naïve plea for science to save us from ourselves. Somehow, our democratic technological society must resolve the dissonance between science and the creation of public policy. Nowhere is this more troublesome than in the formal assessment of risk—the estimation of the association between exposure to a substance and the incidence of some disease, based on scientific data.

Science and the Law at EPA

Here is how the problem emerges at the Environmental Protection Agency. EPA is an instrument of public policy, whose mission is to protect the public health and the environment in the manner laid down by its statutes. That manner is to set standards and enforce them, and our enforcement powers are strong and pervasive. But the standards we set, whether technology- or health-related, must have a sound scientific base.

Science and the law are thus partners at EPA, but uneasy partners. The main reason for the uneasiness lies, I think, in the conflict between the way science really works and the public's thirst for certitude that is written into EPA's laws. Science thrives on uncertainty. The best young scientists flock into fields where great questions have been asked but nothing is known. The greatest triumph of a scientist is the crucial experiment that shatters the certainties of the past and opens up rich new pastures of ignorance.

But EPA's laws often assume, indeed demand, a certainty of protection greater than science can provide with the current state of knowledge. The laws do no more than reflect what the public believes and what it often hears from people with scientific credentials on the 6 o'clock news. The public thinks we know what all the bad pollutants are, precisely what adverse health or environmental effects they cause, how to measure them exactly and control them absolutely. Of course, the public and sometimes the law are wrong, but not all wrong. We do know a great deal about some pollutants and we have controlled them effectively by using the tools of the Clean Air Act and the Clean Water Act. These are the pollutants for which the scientific community can set safe levels and margins of safety for sensitive populations. If this were the case for all pollutants, we could breathe more easily (in both senses of the phrase); but it is not so.

William D. Ruckelshaus is Administrator of the U.S. Environmental Protection Agency, Washington, D.C. 20460. This article is based on a talk he gave at the National Academy of Sciences, Washington, D.C., on 22 June 1983.