

Intercalation, DNA Kinking, and the Control of Transcription

Milton H. Werner, Angela M. Gronenborn,* G. Marius Clore*

Biological processes involved in the control and regulation of transcription are dependent on protein-induced distortions in DNA structure that enhance the recruitment of proteins to their specific DNA targets. This function is often accomplished by accessory factors that bind sequence specifically and locally bend or kink the DNA. The recent determination of the three-dimensional structures of several protein-DNA complexes, involving proteins that perform such architectural tasks, brings to light a common theme of side chain intercalation as a mechanism capable of driving the deformation of the DNA helix. The protein scaffolds orienting the intercalating side chain (or side chains) are structurally diverse, presently comprising four distinct topologies that can accomplish the same task. The intercalating side chain (or side chains), however, is exclusively hydrophobic. Intercalation can either kink or bend the DNA, unstacking one or more adjacent base pairs and locally unwinding the DNA over as much as a full turn of helix. Despite these distortions, the return to B-DNA helical parameters generally occurs within the adjacent half-turns of DNA.

Readout of the information encoded in DNA is often initiated by the assembly of higher order structures comprising multiple proteins bound at distinct DNA sites. The dilemma faced by these proteins is how to shorten the distance between the sites such that the proteins can interact with one another during the process of assembling a nucleoprotein complex. Compacting DNA is a classical problem, first recognized to occur in nuclear chromatin (1) and, more recently, also known to be important in the regulation of transcription (2). The relative stiffness of short segments of the DNA double helix [less than 500 base pairs (bp)] necessitates bending or kinking (3) to bring specified segments into close proximity. The recent determination of several three-dimensional structures of protein-DNA complexes involving such architectural and regulatory proteins has revealed a common theme whereby protein side chain intercalation into the minor groove plays a pivotal role in the mechanism of kinking and bending the DNA (4–8). Protein-induced distortions of DNA structure have long been recognized as an important component of processing information encoded in DNA sequence (3); side chain intercalation, however, as seen in these new structures (4–8), represents one of the most dramatic manipulations of DNA yet observed.

What does side chain intercalation into a DNA helix involve? There are two types of intercalation: partial and com-

plete. In the case of partial intercalation, the intruding side chain unstacks two adjacent base pairs, but does not itself stack over one base pair in lieu of the displaced base pair. In the case of complete side chain intercalation, on the other hand, the side chain is stacked over a base pair in its entirety, fully penetrating to the helix axis whether or not a second base pair is displaced from the DNA helix (9). Complete intercalation is akin to the mode of interaction observed for some DNA binding drugs (10). Whether partial or complete, intercalation unstacks DNA bases at one or more sites, unwinds the DNA, and alters the direction of the helix axis. Twenty years ago, hypothetical structures for the so-called “kinky helix” were proposed to account for the folding of DNA observed under certain conditions, in particular the conformation of DNA in the nucleosome (9). Here we revisit the

kinked helix and discuss the structural parameters of protein-DNA complexes that arise as a direct consequence of protein side chain intercalation.

Minor Groove Intercalation

To date, kinking the DNA by minor groove intercalation of one or more amino acids has only been observed for proteins involved in transcriptional regulation. Specifically, the high-mobility group (HMG) domain proteins SRY (6) and LEF-1 (7), the TATA box-binding protein TBP [*Arabidopsis thaliana*, aTBP; yeast, yTBP (4)], the human *ets1* oncogene product ETS1 (5), and the *Escherichia coli* purine repressor protein PurR (8) use an intercalative wedge to pry open a single base step and distort the DNA (11). Ribbon diagrams of these proteins are shown in Fig. 1. Most strikingly, each protein has a completely different topology, with different parts of the structures being involved in the interactions with DNA.

Distortion of the DNA by partial intercalators is driven by a wedge comprising two to five side chains on the protein surface intercalating into the minor groove (Figs. 2 and 3). The leading edge of the wedge consists of a single hydrophobic amino acid that penetrates within 3 Å (TBP) to 5 Å (SRY, LEF-1, and PurR) of the local helix axis, severely buckling 1 bp (Figs. 2 and 3). There does not appear, however, to be significant stacking of the partially intercalated side chain over either base plane at the site of insertion.

In the case of SRY and LEF-1, the wedge is T-shaped, consisting of four hy-

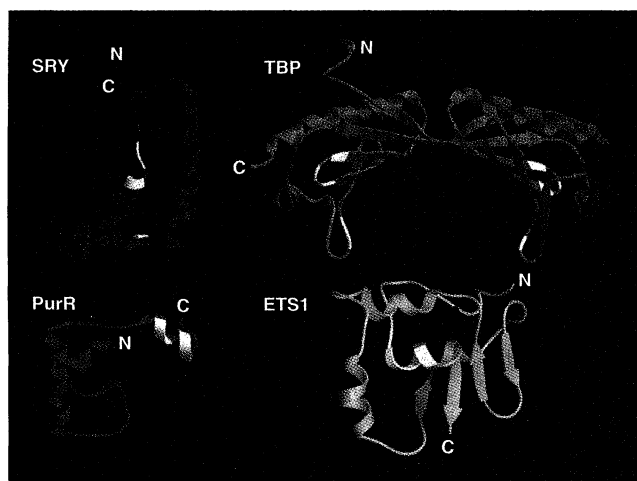


Fig. 1. Three-dimensional structures of minor groove intercalating DNA binding proteins. The folds of SRY, TBP, PurR, and ETS1 are shown as ribbon diagrams (52). For SRY, ETS1, and PurR, only the DNA binding domain is shown. The residues that comprise the intercalative wedge are shown in yellow in each case. The coordinates of SRY, TBP, PurR, and ETS1 are taken from (6), (4), (8), and (5), respectively. N, NH₂-terminus; C, COOH-terminus.

The authors are in Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA.

*To whom correspondence should be addressed.

drophobic amino acids at the crossbar that insert at both the sugar-phosphate backbone and bases and include the partially intercalating side chain at the leading edge (Figs. 2 and 3). The stem of the T consists of an asparagine side chain (Asn¹⁰) that stabilizes the orientation of the wedge by hydrogen bonding primarily to the exocyclic amino group of the guanine 5' to the intercalation site, as well as to the O2 atom of its cytosine partner and to the N3 atom of the adenine in the first A·T base pair at the site of intercalation. Indeed, the ability of the asparagine side chain to form multiple hydrogen bonds is one of the key features distinguishing the specific from the nonspecific DNA binding HMG domains (6). PurR uses a similar T-shaped wedge, but only two (Ala⁵¹ and Leu⁵⁴) instead of four hydrophobic amino acids in each monomer are used (Fig. 3). The wedge in TBP, in contrast, clamps the backbone at the intercalation site by salt bridges to the phosphates rather than by hydrogen bonding to the base step below (Fig. 3). The broader surface of hydrophobic interactions along the protein-DNA interface presumably obviates the need for anchoring the wedge at the neighboring bases as seen in the SRY, LEF-1, and PurR complexes with DNA.

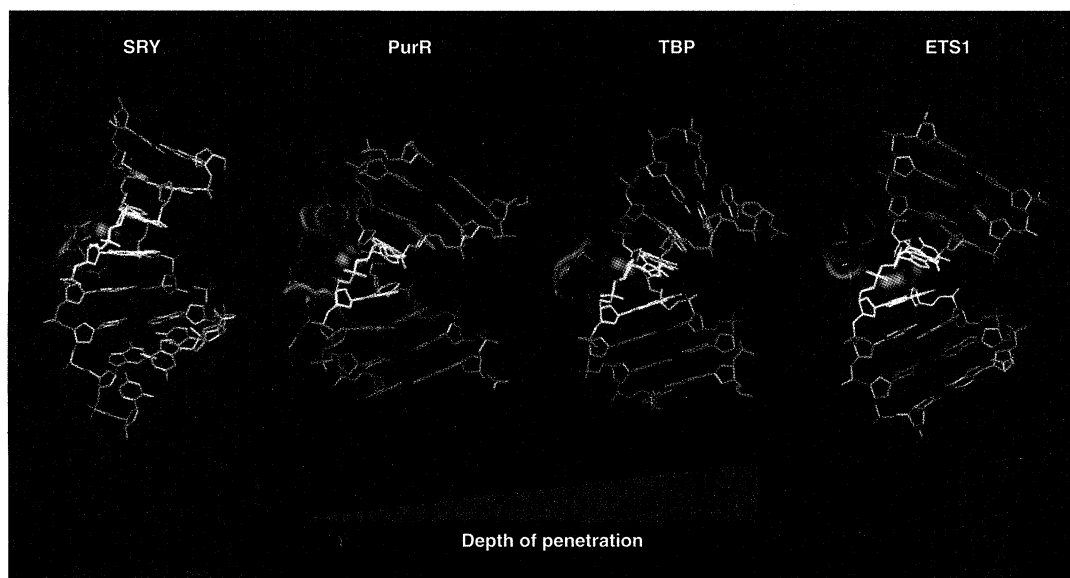
Complete intercalation of a protein side chain, extending all the way from the minor to the major groove, has only been observed in a single case, namely the ETS1-DNA complex (5). The intercalating side chain, Trp²⁸, is positioned on the end of a rigidly held α helix in ETS1 and displaces a C·G base pair with stacking of the invading tryptophan ring over the C·G base pair below (Figs. 2 and 3). Specifically, the six-mem-

bered ring of the side chain of Trp²⁸ is stacked over the base of the cytosine of the 5' C·G base pair. Watson-Crick hydrogen bonding in the displaced C·G base pair 3' to the tryptophan is completely disrupted. Nevertheless, the cytosine and guanine bases of this base pair do not flip out of the DNA helix but are retained in highly propeller-twisted conformations as a result of a number of interactions both in the minor and major grooves (Fig. 4, top). The six- and five-membered rings of Trp²⁸ are approximately orthogonal to the bases of the displaced cytosine and guanine bases, respectively, thereby providing weak hydrogen bonding interactions in which the protons of the tryptophan ring act as acceptors and the π electrons of the displaced bases, as donors. In the major groove, the hydroxyl group of Tyr⁸⁶ forms a hydrogen bond to the phosphate of the displaced cytosine, and the aromatic ring of Tyr⁸⁶ is simultaneously stacked against the displaced cytosine and lies approximately orthogonal to the guanine ring 3' to the disrupted base pair. Thus, Tyr⁸⁶ acts like a dam to hold the disrupted cytosine in place (Fig. 4, top). The intercalative complex is further stabilized on the minor groove side by hydrophobic interactions between the methyl groups of Leu²⁷ and Leu³¹ and the displaced base pair, and by an electrostatic interaction from the side chain of Gln²⁶ to the exocyclic amino group of the guanine of the C·G base pair 5' to the tryptophan and to the phosphate of the G·C base pair 5' to the intercalation site.

The conformation of proteins presently known to intercalate a side chain (or side chains) into DNA exhibit no common structural motif with regard to their DNA

binding surfaces (Fig. 1). LEF-1 and SRY are essentially helical proteins that present a boomerang- or L-shaped concave binding surface to the DNA. TBP is saddle-shaped with a DNA binding surface formed by a 10-stranded antiparallel β sheet. PurR and ETS1 have bipartite DNA binding surfaces with a helix-turn-helix major groove recognition motif and an additional helix (which in the case of PurR is also part of the dimerization interface) that presents the intercalating wedge to the adjacent minor groove (Fig. 1). In general, the protein topology at the DNA binding surface is complementary to that of the distorted DNA structure, with only modest changes in protein conformation upon binding to DNA. On this basis, the protein binding surface represents the driving force for induced fit of the DNA conformation during the process of molecular recognition of the DNA sequence at the site of insertion. Upon binding DNA, there is a substantial change in accessible surface area for these proteins, with more than 1000 Å² of surface buried, ~70% of which is hydrophobic in the case of TBP, SRY, and LEF-1. For PurR and ETS1, the minor groove interaction surface in the vicinity of the intercalating side chain is predominantly hydrophobic, but the balance of the DNA binding surface is both hydrophobic and hydrophilic. The protein and DNA surfaces at the minor groove binding interface are packed tightly with nonpolar amino acids lining the floor of the interaction surface, whereas polar amino acids are distributed either at the edges or as bumps in the floor, poised to interact with either the phosphates or the base nitrogens or oxygens, respectively.

Fig. 2. Depth of penetration into the DNA helix by minor groove intercalating proteins. The depth of penetration for the minor groove intercalating wedge increases from left to right, and the blue scale serves solely to guide the eye. The intercalating side chains at the leading edge of the wedge are shown in red as a molecular surface and are Ile¹³ for SRY, Leu⁵⁴ and the symmetrically related Leu^{54'} for PurR, Phe⁵⁷ for aTBP (only one of two shown for aTBP), and Trp²⁸ for ETS1. (The other intercalating side chain for aTBP is Phe¹⁴⁸, and for yTBP, they are Phe⁹⁹ and Phe¹⁹⁰.) The additional amino acid side chains that form the remainder of the intercalating wedge are shown in magenta, with the relevant portion of the protein backbone in green. The base step at which intercalation takes place is indicated in yellow (see Table 1). The graphics for this figure, as well as for Figs. 3, 4, and 6, were generated with the program GRASP (53).



Distortion of DNA Structure upon Minor Groove Intercalation

The helical structure of the DNA is altered significantly as a consequence of side chain intercalation. In order to accommodate the intercalating side chain, the minor groove is widened, reducing the helical twist locally and introducing a significant positive roll at one or more base steps (11). The definitions of the various terms used to describe the DNA structure are illustrated schematically in Fig. 5, and ribbon diagram representations of the DNA in the complexes with SRY, TBP, PurR, and ETS1 are shown in Fig. 6.

The extent of unwinding varies between 5° (for the ETS1-DNA complex) and 20° (for the TBP-DNA complex) at the intercalation site, with a positive roll of 20° to 50° introduced at the disrupted base step (Table 1). To accommodate the intercalating side chain, the helical rise of the DNA increases to 4 Å in the SRY-DNA complex and up to 6 to 7 Å in the

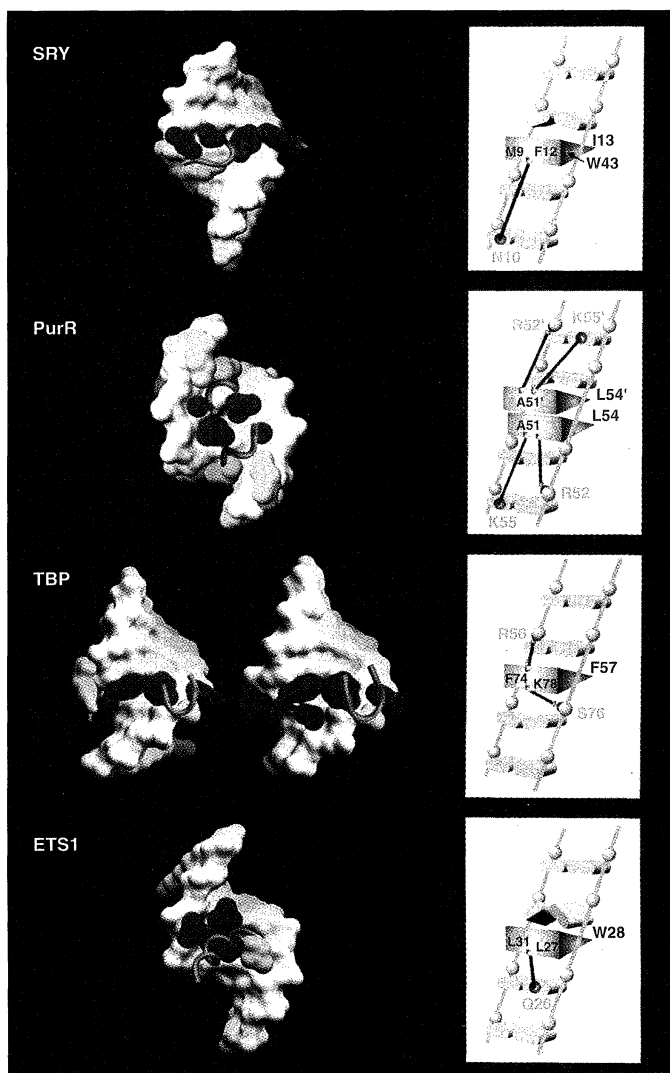
case of the PurR-DNA and ETS1-DNA complexes. The larger helical rise is correlated with the depth of penetration by the leading edge of the wedge (Fig. 2). The consequence of intercalation by ETS1 and PurR is a kink in the DNA (5, 8). Kinking is the result of unstacking a single base step and rolling open the base planes to introduce a sharp, localized bend (9, 12). SRY, LEF-1, and TBP extend their kinks along the length of the interaction surface by partial unstacking and rolling of several additional base steps, distributing the energy penalty for altering the DNA conformation over several base pairs within one helical turn. By kinking and rolling the DNA, SRY, LEF-1, and TBP bend the helix over a full turn and introduce a significant lateral displacement between the incoming and outgoing helix axis (for example, 18 Å in the case of the TBP-DNA complex). The recovery to B-DNA geometry is rapid in the TBP-DNA and PurR-DNA complexes, which return to B-like helical twist, rise, and minor groove

width within two base steps either 5' or 3' to the intercalation site; in the SRY-DNA and ETS1-DNA complexes, the recovery of the DNA to B-helical parameters occurs within five base steps.

The predilection of a particular DNA sequence toward specific helical distortions has been the subject of many experimental (13) and theoretical analyses (14) that indicate that flexibility and bendability contribute significantly to the ability of a protein to bind to a specific target. Studies of nucleosomal DNA (15) and DNA bound by the catabolite gene activator protein CAP (16) suggest that compression of the minor groove is more favorable in A·T-rich sequences, whereas G·C-rich sequences tend to distort with major groove compression. Indeed, ETS1 and PurR narrow the major groove at CpC and CpG steps, respectively (Table 1). In TBP, SRY, and LEF-1, however, DNA binding at A·T-rich sequences widens the minor groove and helically unwinds the DNA, driving the helix geometry toward an alternative conformation that is more akin to an A-type helix (17). It is now clear that the flexibility of DNA is dependent not only on the base composition at a target site (18) but also on the sequences in regions flanking those sites (19). Analysis of dinucleotide geometries from the database of B-DNA crystal structures begins to shed light on the choice of target sequence at which intercalation takes place. CpG, GpG (that is, CpC), ApG, and, to a lesser extent, ApA steps have lower average twist relative to other dinucleotide steps (20). Moreover, the helical twist at dinucleotide steps is inversely related to the roll angle, that is, a lower average twist is correlated with an increased positive roll (20). It is possible, therefore, that proteins that alter DNA structure through minor groove intercalation target certain sequences, at least in part, based on this inherent tendency of certain dinucleotide steps to underwind and roll. These effects may be additive in A·T-rich segments because TBP, SRY, and LEF-1 prefer targets with a string of 3 to 5 bp of A·T-rich sequence and underwind the DNA over nearly a full turn of helix. One cannot, however, fully account for target identification by minor groove intercalators on the basis of dinucleotide geometries alone. TpA is targeted by TBP; yet, on average, it is the least underwound in B-DNA crystal structures (20). TpG in SRY is not as rolled to open the minor groove as is TpT (that is, ApA), even though TpG can potentially occupy the broadest range of conformational space (21).

There are distinct chemical differences in the minor groove surfaces of A·T versus G·C base pairs (22). A·T base pairs have hydrogen bond acceptors in the minor

Fig. 3. Intercalating wedges of minor groove DNA binding proteins. The molecular surface of each protein-DNA complex is shown in gray scale for DNA, red for the hydrophobic protein side chains, and yellow for the protein residues that act as electrostatic anchors. The wedge is T-shaped for SRY [from left to right: Met⁹, Ile¹³, Phe¹², Trp⁴³, and Asn¹⁰ (yellow)], LEF-1 (not shown) (50), and PurR [from left to right: Leu⁵⁴, Lys⁵⁵ (yellow), Ala⁵¹, Arg⁵² (yellow), and their symmetry-related partners (above)]. Two examples of the wedge from aTBP: the side chains in the left molecule (from left to right) are Arg⁵⁶ (yellow), Phe⁵⁷, Phe⁷⁴, Ser⁷⁶ (yellow), and Lys⁷⁸; the side chains in the right molecule (from left to right) are Leu¹⁴⁵, Phe¹⁴⁸, Pro¹⁴⁹, Phe¹⁶⁵, Ser¹⁶⁷ (yellow), and Lys¹⁶⁹. In ETS1, the intercalative wedge is formed (from left to right) by Leu³¹, Leu²⁷, Trp²⁸, and Gln²⁶ (yellow). To the right in each figure is a mechanical model depicting in a schematic manner the mode of side chain intercalation in each case. In the case of TBP only the left model is schematized.



groove at the N3 and O2 positions of adenine and thymine, respectively. G·C base pairs, in contrast, have both hydrogen bond donors (the 2-amino group of guanine) and hydrogen bond acceptors (O2 and N3 atoms of cytosine and guanine, respectively). Thus, a G·C base pair may be distinguished from an A·T base pair on the basis of its relative ability to form certain types of hydrogen bonds. A G·C base pair may also be distinguished from an A·T one on the basis of steric clash in the minor groove because the A·T base pair lacks the bulky exocyclic amino group of a guanine. Indeed, some minor groove binding drugs identify G·C base pairs by hydrogen bond formation to the exocyclic amino group of guanine (10). By steric avoidance of this group, drugs such as thiotropin choose A·T base pairs (23). Hydrogen bond formation to the exocyclic amino group of a guanine has also been suggested to influence the anisotropic flexibility and bendability of DNA, which could in turn significantly influence the choice of target by an intercalating DNA binding protein whose chief objective is to kink or bend DNA (24). Yet, SRY and LEF-1 use the same amino acid, namely a tyrosine, albeit at a slightly different position in the sequence, in very similar conformational arrangements (6, 7) to distinguish two A·T base pairs in an SRY target sequence (AACAAA) from two T·A base pairs in a LEF target sequence (TTCAAA), respectively (25), even though these base pairs look virtually identical in the minor groove with respect to their ability to form certain types of directional interactions (22). This arrangement leaves the more subtle hydrophobic effects in the minor groove as the distinguishing chemical principle. Thus, for example, the surface of adenosine in the minor groove is more hydrophobic than that of a guanine, owing to the presence of a proton in the 2-position in the case of the former and its replacement by an amino group in the case of the latter.

Protein-DNA Interactions in the Minor Groove

The interaction of minor groove intercalating proteins with DNA is primarily hydrophobic at the site of insertion. For TBP, SRY, and LEF-1, these nonpolar interactions extend over most of the interaction surface between the protein and the DNA. Base recognition occurs through van der Waals interactions between β - or γ -branched amino acids (Val, Ile, or Leu) or the faces of aromatic amino acids (Phe or Tyr) and the minor groove edges and faces of the DNA bases. Relatively few hydrogen bonds between protein side chains and the DNA bases are present at the interface, with Ser (SRY), Thr (TBP), Asn (SRY, LEF-1, and TBP), or Tyr (SRY) forming base-specific

hydrogen bonds with either the adenine-N3, thymine-O2 atoms, or the exocyclic amino group of guanine. There appears to be no correlated pattern of hydrogen bond formation between a DNA base and a particular amino acid. No water-mediated contacts with DNA bases are observed in the aTBP, yTBP (4), SRY (26), PurR (8), or ETS1 (26) complexes, although water appears to play a role in stabilizing a compressed major groove in the aTBP complex (4). SRY, LEF-1, and TBP display extensive interactions with the sugar-phosphate backbone (4, 6, 7). Salt bridges or hydrogen bonds form between DNA phosphates and Lys or Arg residues of the protein at nearly every phosphate position at the protein-DNA interface, which

includes water-mediated salt bridges with several phosphates in the case of aTBP (4). ETS1 (5) and PurR (8) form less extensive contacts with the backbone of the DNA, with salt bridges formed at only one out of four possible phosphates at the intercalation site. Despite the numerous electrostatic contacts with the DNA backbone, complexes of SRY and ETS1 are relatively resistant to titration with monovalent salts and appear to be stable to NaCl or KCl concentrations greater than 0.8 M (27, 28). This property emphasizes the importance of hydrophobic interactions in stabilizing these complexes.

The DNA distortions induced by minor groove intercalating proteins suggest a complex mechanism of sequence recognition.

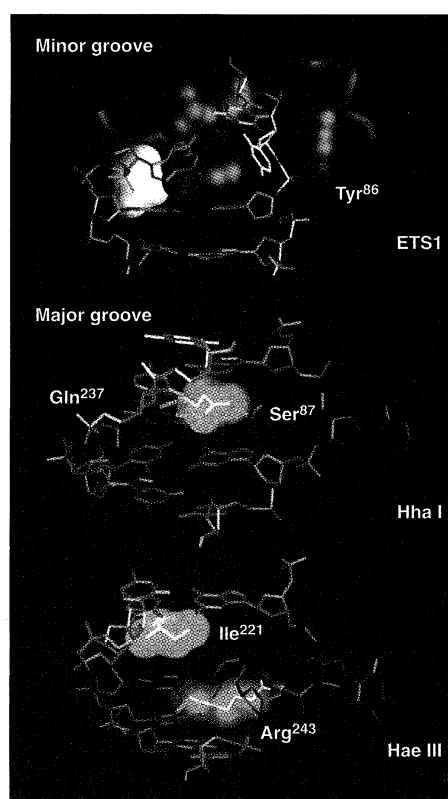


Fig. 4. Stabilization of base pair disruption by intercalative DNA binding proteins in the minor (top) and major (bottom) grooves. Stabilization of the disrupted base pair in the ETS1-DNA complex (top) is achieved through stacking interactions with Tyr⁸⁶ that prevent the expulsion of Cyt7 (yellow) from the helix. The guanine partner of Cyt7 (Gua28, magenta) is stabilized through interactions with the intercalating Trp²⁸ (red, center). The remainder of the wedge is formed by Leu²⁷ and Leu³¹ (red, top, left to right) and Gln²⁶ (yellow, left). In the case of the major groove intercalating methyltransferases (middle and bottom), Hha I (37) stabilizes the orphan guanine (red) by electrostatic interactions (blue hashed lines) with Gln²³⁷ (yellow in transparent gray surface), which inserts from the major groove and partially stacks over the guanine 3' to the site of disruption. Ser⁸⁷ (magenta) approaches from the minor groove and reinforces these contacts by hydrogen bond formation with the carboxamide oxygen of Gln²³⁷. Hae III (38) stabilizes the guanine opposite the flipped-out cytosine (red) by re-pairing with a cytosine from the 3' base pair (magenta). The orphan guanine (magenta), now 3' to the flipped-out cytosine, is stabilized by electrostatic interactions (blue hashed lines) with Arg²⁴³ (lower yellow in transparent gray surface). The hole left by re-pairing bases is partially filled from the major groove by stacking of Ile²²¹ (upper yellow in transparent gray surface) over the guanine partner (red) of the flipped-out base.

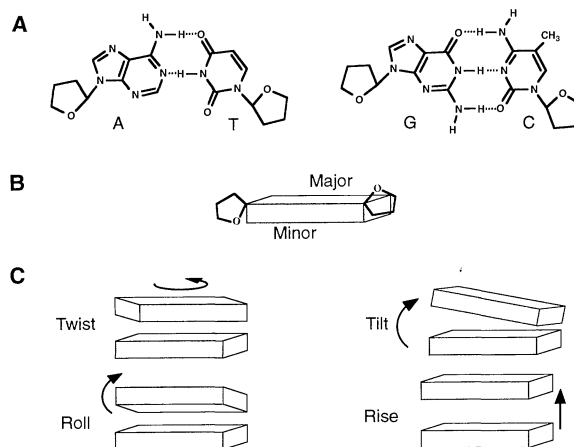


Fig. 5. Schematic illustration of the various helical parameters used to describe DNA structure. (A) Watson-Crick base pairing in DNA. (B) Projection of a base pair as a three-dimensional box viewed along the minor groove of the base pair (shaded) (54). (C) Geometric definitions of twist, roll, tilt, and rise discussed in the text. The positive sense of rotation is indicated in each case.

Although only limited data are available, binding kinetics for the interaction of the ETS1 DNA binding domain (29) and γ TBP (30–33) with their DNA targets indicate that the apparent association rate constant, k_{on} , for these proteins is on the order of 0.3×10^6 to $1.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, considerably slower than the diffusion-limited on-rate often cited for DNA binding proteins (34). Off-rates for these proteins are relatively slow, with half-lives ranging from about 3 min for the ETS1 DNA binding domain (29) to about 60 min (depending on monovalent salt concentration) for γ TBP (30). Analysis of the binding kinetics for γ TBP (30) and the full-length ETS1 isoforms p51 and p42 (28), however, indicate that different pathways can be followed to reach the final intercalated protein-DNA complex. Kinetic modeling for the ETS1 isoforms suggests the presence of an intermediate state that forms rapidly but is slowly converted to the final species (28). The lifetime of this intermediate state varies for the ETS1 isoforms, with half-lives ranging from 5 s for p51 to 100 s for p42 (28). In contrast, TBP kinetics appear to be purely second order, with no hint of a stable intermediate being formed on the pathway to a final complex (31).

Intuitively, the distortion of the DNA should pose the greatest barrier to complex formation, particularly for TBP, SRY, and

LEF-1, which helically unwind the DNA. If bending poses the greatest energy barrier, then a pre-bent sequence should have faster association kinetics, a phenomenon observed for γ TBP (33) but not yet tested for ETS1. Moreover, the extent of DNA bending appears to be directly proportional to the lifetime of the TBP-DNA complexes, that is, greater affinity parallels increased bending for a given site (35). Thus, binding and bending do not appear to be thermodynamically separable for TBP (31). In contrast, DNA binding studies of a mutant of SRY [Met⁹→Ile (36)] and studies of LEF-1 binding to a mutated target site with an A→T transversion [TTCAATG (7)] suggest that high-affinity binding can occur with substantially reduced DNA bending. This observation has led to the suggestion that the free energy of complex formation can somehow be partitioned between binding affinity and DNA bending for the HMG domain proteins (36). A simpler explanation may be that alternative contacts form in these mutant complexes that are compensatory with respect to the free energy of binding but fail to drive the DNA distortion to completion. Only detailed structural analysis of these mutant complexes will permit a clear understanding of the linkage between sequence specificity and the DNA conformation that is induced upon binding.

An Alternative Mode of Intercalation into the Major Groove

The structures of two methyltransferase-DNA complexes indicate that an alternative mode of intercalation is possible in different functional environments [Hha I methyltransferase (37) and Hae III methyltransferase (38)]. These enzymes completely “flip out” a cytosine base for modification by methylation. The hole left in the DNA helix is partly filled by protein side chains that approach the DNA from the major groove and stabilize the orphaned base through hydrogen bonding and van der Waals interactions (Fig. 4, middle and bottom). Gln²³⁷ in Hha I methyltransferase stabilizes the guanine that has lost its Watson-Crick partner by hydrogen bonding the carboxamide oxygen and main-chain amide hydrogen to the guanine N1, N2, and O6 atoms. The glutamine is deeply inserted into the DNA helix and stacks over the guanine 3' to the extrahelical cytosine without additional rearrangement of the neighboring base pairs in the target sequence. In the case of the Hae III methyltransferase-DNA complex, cytosine expulsion occurs with base pair rearrangements in the GGCC recognition element. In this instance, the orphaned guanine is

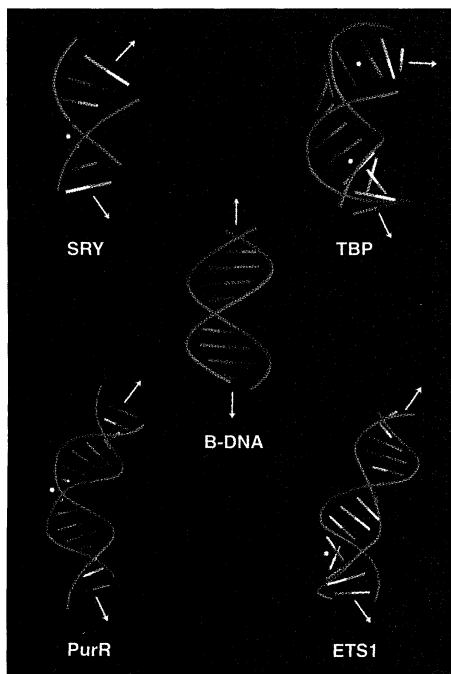


Fig. 6. Schematic ribbon diagrams of the DNA bound to SRY, TBP, PurR, and ETS1 compared to canonical B-DNA. The approximate directions of the helix axes are indicated by arrows. A-T base pairs are shown as blue and red, G-C base pairs as green and yellow, and the sugar-phosphate backbone as blue and red ribbons. The ivory dot indicates the site of wedge intercalation in each case.

Table 1. DNA helix parameters at the site of minor groove intercalation. Helix parameters were calculated with the program CURVES (48) or CompDNA (49) for the following DNA sequences: SRY (dG₁CACAAAC₈dG₉TTTGTGC₁₆) (50), PurR (dA₁CGAAAAACGTTTCGT₁₅dA₁₆CGAAAAACGTTTCGT₃₂), aTBP (dG₁CTATAAAAGGGCA₁₄dT₁₅GCCCTTTTATAGC₂₈), γ TBP (dG₁TATATAAACGG₁₃dG₁₄CGTTTATATAC₂₆), and ETS1 (dT₁CGAGGCCGGAAGTTTCA₁₇dT₁₈CGAACTTCGGCTCGA₃₄). The parameters are listed in order of increasing depth of penetration into the DNA helix by the leading edge of the wedge (aTBP and γ TBP are equally penetrant). The base step at the intercalation site is indicated in bold type. Parameters for B-DNA or A-DNA are calculated from a canonical helix derived from the database of B-DNA or A-DNA crystal structures (49). The values for the helical parameters are reported as averages of the calculations from the two programs, which varied no more than 3° for twist, 2° for roll, and 3° (in the case of the ETS1-DNA complex only) for tilt. The calculation of rise varies widely and is reported as CURVES/CompDNA measurements. Groove widths are reported relative to the shortest distance in canonical B-DNA, $O4'_{(i)} - O4'_{(i'+4)}$ in the minor groove and $P_{(i)} - P_{(i'+4)}$ in the major groove, where i is the nucleotide on the sense strand and i' is the nucleotide on the antisense strand, less twice the van der Waals radius of oxygen (1.4 Å) or phosphorus (2.9 Å), respectively. These distances are $O4'_{(i)} - O4'_{(i'+2)}$ and $P_{(i)} - P_{(i'+5)}$ in A-DNA. For the distorted DNAs, the shortest distance observed in the minor or major groove at the center of the intercalated base step is reported in parentheses as calculated by the program CURVES (48, 51). The kink angle is calculated between the incoming and outgoing cylindrical axes of the B-like portion of the distorted DNA structures; for SRY, this is determined from model building as a result of the shortness of the DNA duplex studied (6).

Protein	Step	Twist (°)	Roll (°)	Tilt (°)	Rise (Å)	Groove width		Kink (°)
						Minor groove (Å)	Major groove (Å)	
SRY	ApA	23	21	1	4.3/3.9	12.8 (9.8)	11.1 (10.6)	65
PurR	CpG	26	49	0	7.0/5.8	12.7 (8.9)	11.2 (6.8)	45
aTBP	TpA	15	47	−8	5.8/4.7	12.0 (7.3)	10.7 (7.6)	45
	ApG	26	43	3	5.9/5.0	11.8 (7.8)	9.4 (7.6)	45
γ TBP	TpA	21	46	−0.8	5.3/4.3	11.8 (7.5)	11.4 (11.4)	45
	ApA	18	38	2	5.6/4.7	11.7 (8.2)	10.9 (9.4)	45
ETS1	CpC	31	48	−11	7.2/5.9	13.0 (10.2)	5.4 (3.9)	60
B-DNA		36	0.2	0	3.4	4.5	11.5	
A-DNA		31	7.5	0	3.3	6.0	2.5	

not the partner of the extrahelical cytosine, rather it is the guanine from the base pair 3' to the site of expulsion. The guanine partner of the extrahelical cytosine re-pairs with the 3' cytosine in a highly propeller-twisted Watson-Crick fashion, stabilized by partial major groove intercalation and stacking of the γ and δ methyl groups of Ile²²¹ over the guanosine ring. The unpaired guanine is stabilized by hydrogen bonding interactions between the N7 atom of the base and the guanidinium group of Arg²⁴³. Thus, despite the different roles of intercalation in DNA modification and transcriptional regulation, structural parallels exist in the manner in which distorted DNA structures are stabilized.

Structural Determinants of Biological Function

Numerous biological functions that use sequence-encoded information in DNA are dependent on distally bound proteins often several kilobases away from the point of influence (3). The idea that a cis-acting element can simply be a structural determinant was first demonstrated for *E. coli* integration host factor (IHF), a protein that promotes the assembly of the bacteriophage λ intasome through protein-induced DNA bending (39). The ability to functionally replace an IHF-induced bend with a pre-bent DNA sequence or an appropriately phased bend induced by an alternative protein established that architectural elements play a critical role in nucleoprotein assembly (40). Preliminary crystallographic evidence suggests that IHF partially inserts a proline at an ApA step in the minor groove (41), placing IHF in the category of minor groove intercalating proteins.

Understanding how higher order structures assemble requires knowledge of the nature of the protein-protein interactions that inevitably form between different components when DNA is locally bent by an architectural protein. The interactions between proteins bound to distinct DNA targets that additionally interact with each other seem to be mediated by acidic domains that by themselves are poorly structured (42) but may be capable in certain cases of forming elements of regular secondary structure under conditions that may mimic the environment of a condensed state in a nucleoprotein assembly. Thus, it appears that a cooperative assembly pathway exists in which architectural proteins bend DNA, bringing distantly bound proteins into contact, thereby stimulating protein-protein interactions to complete the process. Structures of two such protein-protein interaction domains from proteins involved in the basal transcription machinery (2), one from TFIIB (43) and the other

from TFIIS (44), indicate that these two domains contain a common structural element, namely a zinc ribbon. The first insight into the condensed state of a nucleoprotein assembly has been revealed from the structure of the core protein binding domain of TFIIB bound to a TBP-DNA complex (45). This structure provides the first picture not only of the protein-protein interactions associated with an architectural protein (TBP) when bound to DNA, but also of the additional role DNA bending can play in positioning target sequences such that they interact with more than one protein simultaneously (45).

The critical importance of DNA binding architectural proteins in the control of transcription is apparent from the consequence of a loss of regulatory control in mammalian sex determination. Fifteen percent of known cases of 46X,Y sex reversal in humans are the result of a failure of the SRY protein to switch on the development of the testes from the bipotential embryonic gonad, resulting in phenotypic females with a male karyotype. In all but one case, the loss of SRY function occurs as a result of point mutations or deletions in the DNA binding domain of the protein that directly disrupt the ability of SRY to promote the assembly of a higher order structure (6, 46). These include a mutation of the intercalating side chain Ile¹³ to a Thr, and a mutation of Met⁹, which is part of the hydrophobic intercalative wedge, to Ile. Hence the architectural components of the transcription machinery are just as critical to function as the enzymological ones.

DNA bending has long been recognized as an important component of biological activity. Intercalation represents only one mechanism by which protein-induced DNA bending can occur (47). Many questions remain unanswered, in particular the ill-understood nature of sequence recognition in the minor groove. While some of the pertinent principles are beginning to be unraveled based on the structures described above, the relative lack of directional contacts (that is, salt bridges or hydrogen bonds) leaves open the question of how the sequence is initially located and recognized. Nevertheless, some common features have emerged from the structures of these intercalative proteins, but as is frequently the case in matters of molecular recognition, there is a wide diversity of architectural frameworks upon which a particular recognition theme may be grafted; nature has numerous ways to evolve functionally related activities.

REFERENCES AND NOTES

1. R. D. Kornberg, *Science* **184**, 868 (1975); A. L. Olins and D. E. Olins, *ibid.* **183**, 330 (1974); K. E. van Holde, B. Sahasrabudhe, R. Shaw, *Nucleic Acids Res.* **1**, 1579 (1974); J. D. Griffith, *Science* **187**, 1202 (1975); J. P. Baldwin, P. G. Boseley, E. M. Bradbury, K. Ibel, *Nature* **253**, 245 (1975).
2. R. Tjian and T. Maniatis, *Cell* **77**, 5 (1994).
3. J. C. Wang and G. N. Glaever, *Science* **240**, 300 (1988); A. A. Travers, in *DNA-Protein: Structural Interactions*, D. M. J. Lilley, Ed. (Oxford Univ. Press, New York, 1995), pp. 49–75.
4. J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **365**, 520 (1993); J. L. Kim and S. K. Burley, *Nature Struct. Biol.* **1**, 638 (1994); Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **365**, 512 (1993).
5. M. H. Werner *et al.*, *Cell* **83**, 761 (1995).
6. M. H. Werner, J. R. Huth, A. M. Gronenborn, G. M. Clore, *ibid.* **81**, 705 (1995).
7. J. J. Love *et al.*, *Nature* **376**, 791 (1995).
8. M. A. Schumacher, K. Y. Choi, H. Zalkin, R. G. Brennan, *Science* **266**, 763 (1994).
9. J. A. Schellman, *Biopolymers* **13**, 217 (1974); F. H. C. Crick and A. Klug, *Nature* **255**, 530 (1975); H. M. Sobell, C.-C. Tsai, S. G. Gilbert, S. C. Jain, T. D. Sakore, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3068 (1976). With respect to the definitions put forth in this article, kinking of the DNA is distinguished from bending. Kinking occurs when 2 bp are unstacked, resulting in an abrupt, local alteration in the direction of the helix axis. Bending refers to a more general phenomenon in which a change in the direction of the helix axis can occur with or without unstacking of adjacent base pairs. Thus, kinking is one mode of bending, but bending can occur without kinking.
10. M. S. Searle, *Prog. Nucl. Magn. Reson. Spectrosc.* **25**, 403 (1993); B. H. Geierstanger and D. E. Wemmer, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 463 (1995).
11. A base step is defined as two consecutive base pairs of DNA. Base pairs are formed from four different building blocks as illustrated in Fig. 5A. Two of these building blocks hydrogen bond to form a base pair which, when viewed from the minor or major groove edges (Fig. 5B), can be thought of as comprising a single plane. Hence, geometrically, a base pair can be treated as a single plane—a base plane. When one base plane is tilted about its long axis relative to an adjacent base pair, the angle of tilting is defined as roll. Thus, base planes may be "rolled open" relative to each other either toward the minor or major grooves of DNA (Fig. 5C). Unwinding of the DNA is accomplished by lowering the degree of twist (Fig. 5C).
12. V. B. Zhurkin *et al.*, *Nucleic Acids Res.* **6**, 1081 (1979); P. J. Hagerman, *Biochem. Biophys. Acta* **1131**, 125 (1992).
13. A. A. Travers, *Annu. Rev. Biochem.* **58**, 427 (1989); D. M. Crothers, T. E. Haran, J. G. Nadeau, *J. Biol. Chem.* **265**, 7093 (1990); T. E. Haran, J. D. Kahn, D. M. Crothers, *J. Mol. Biol.* **244**, 135 (1994); A. A. Travers, *Nature Struct. Biol.* **2**, 615 (1995).
14. L. W. Keepers, P. A. Kollman, P. K. Weiner, T. L. James, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5537 (1982); M. Levitt, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 251 (1982); S. J. Weiner, *et al.*, *J. Am. Chem. Soc.* **106**, 765 (1984); K. K. Irikura, B. Tidor, B. R. Brooks, M. Karplus, *Science* **229**, 571 (1985); C. Singh, P. Weiner, P. A. Kollman, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 755 (1985); S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comp. Chem.* **7**, 230 (1986); S. N. Rao, U. C. Singh, P. A. Kollman, *Isr. J. Chem.* **27**, 189 (1986).
15. H. R. Drew and A. A. Travers, *J. Mol. Biol.* **186**, 773 (1985); S. C. Satchwell, H. R. Drew, A. A. Travers, *ibid.* **191**, 659 (1986); C. R. Calladine and H. R. Drew, *ibid.* **192**, 907 (1986).
16. M. R. Gartenberg and D. M. Crothers, *Nature* **333**, 824 (1988); D. D. Dalma-Weiszhausz, M. R. Gartenberg, D. M. Crothers, *Nucleic Acids Res.* **19**, 611 (1991).
17. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
18. M. E. Hogan, M. W. Roberson, R. H. Austin, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9273 (1989); J. D. Kahn, E. Yun, D. M. Crothers, *Nature* **368**, 163 (1994).
19. G. B. Koudelka, P. Harbury, S. C. Harrison, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4633 (1988); M. Lundin, J. O. Nehlin, H. Ronne, *Mol. Cell. Biol.* **14**, 1979 (1994).
20. A. A. Gorin, V. B. Zhurkin, W. K. Olson, *J. Mol. Biol.* **247**, 34 (1994).

21. S. Cheung, K. Arndt, P. Lu, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3665 (1984); D. Battacharya and M. Bansal, *J. Biomol. Struct. Dyn.* **8**, 539 (1990); K. Yanagi, G. G. Privé, R. E. Dickerson, *J. Mol. Biol.* **217**, 201 (1991); E. Hassan, thesis, University of Cambridge (1995).
22. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
23. S. Kumar et al., *J. Biomol. Struct. Dyn.* **8**, 99 (1990).
24. C. Bailly, M. J. Waring, A. A. Travers, *J. Mol. Biol.* **253**, 1 (1995).
25. K. Giese, J. Cox, R. Grosschedl, *Cell* **69**, 185 (1992); V. R. Harley, R. Lovell-Badge, P. N. Goodfellow, *Nucleic Acids Res.* **22**, 1500 (1994).
26. M. H. Werner, A. M. Gronenborn, G. M. Clore, unpublished results.
27. M. H. Werner, R. J. Fisher, G. M. Clore, A. M. Gronenborn, unpublished results.
28. R. J. Fisher et al., *Protein Sci.* **3**, 257 (1994).
29. R. J. Fisher, M. Fivash, S. V. Bladen, unpublished results.
30. B. C. Hoopes, J. F. LeBlanc, D. K. Hawley, *J. Biol. Chem.* **267**, 11539 (1992).
31. V. Petri, M. Hsieh, M. Brenowitz, *Biochemistry* **34**, 9977 (1995); D. Sun and L. H. Hurley, *Chem. Biol.* **2**, 457 (1995); K. M. Parkhurst, M. Brenowitz, L. J. Parkhurst, personal communication.
32. G. M. Perez-Howard, P. A. Weil, J. M. Beechem, *Biochemistry* **34**, 8005 (1995).
33. J. D. Parvin, R. J. McCormick, P. A. Sharp, D. E. Fisher, *Nature* **373**, 724 (1995).
34. M. T. Record, J.-H. Ha, M. A. Fisher, *Methods Enzymol.* **208**, 291 (1991).
35. D. B. Starr, B. C. Hoopes, D. K. Hawley, *J. Mol. Biol.* **250**, 434 (1995).
36. A. Pontiggia et al., *EMBO J.* **13**, 6115 (1994).
37. S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **76**, 357 (1994); X. Cheng, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293 (1995); X. Cheng, *Curr. Opin. Struct. Biol.* **5**, 4 (1995); S. Mi, D. Alonso, R. J. Roberts, *Nucleic Acids Res.* **23**, 620 (1995).
38. K. M. Reinisch, L. Chen, G. L. Verdine, W. N. Lipscomb, *Cell* **82**, 143 (1995).
39. L. Moitoso de Vargas, S. Kim, A. Landy, *Science* **244**, 1457 (1989); H. A. Nash, *Trends Biochem. Sci.* **15**, 222 (1990); H. A. Nash, in *Regulation of Gene Expression in Escherichia coli*, E. C. C. Lin and A. S. Lynch, Eds., in press. Cis-acting element refers to a protein whose actions are felt at the same DNA locus at which it binds.
40. S. D. Goodman and H. A. Nash, *Nature* **341**, 251 (1989); A. M. Segall, S. D. Goodman, H. A. Nash, *EMBO J.* **13**, 4536 (1994).
41. P. Rice, S.-W. Yang, K. Mizuuchi, H. Nash, personal communication.
42. E. Giniger and M. Ptashne, *Nature* **330**, 670 (1987); P. B. Sigler, *ibid.* **333**, 210 (1988); L. Donaldson and J. P. Capone, *J. Biol. Chem.* **267**, 1411 (1992); P. O'Hare and G. Williams, *Biochemistry* **31**, 4150 (1992).
43. W. Zhu et al., *Nature Struct. Biol.*, in press.
44. X. Quan et al., *Biochemistry* **32**, 9944 (1993).
45. D. B. Nikolov et al., *Nature* **377**, 119 (1995).
46. P. N. Goodfellow and R. Lovell-Badge, *Annu. Rev. Genet.* **27**, 71 (1993); K. McElreavey, E. Vilain, C. Cotinot, E. Payen, M. Fellous, *Eur. J. Biochem.* **218**, 769 (1993).
47. Bending without intercalation or exclusive minor groove binding occurs in the protein-DNA complexes of several transcription factors from prokaryotes and eukaryotes, for example, *E. coli* CAP protein [S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **253**, 101 (1991)], HNF3/forkhead [K. L. Clark, E. D. Halay, E. Lai, S. K. Burley, *Nature* **364**, 412 (1993)], and the $\alpha 1/\alpha 2$ homeodomain [T. Li, M. R. Stark, A. D. Johnson, C. Wolberger, *Science* **270**, 262 (1995)]. Kinking occurs in several DNA enzymes without intercalation, for example, deoxyribonuclease I [S. A. Weston, A. Lahm, D. Suck, *J. Mol. Biol.* **226**, 1237 (1992)], Eco RV [F. K. Winkler et al., *EMBO J.* **12**, 1781 (1993)], and Eco RI [J. A. McClarin et al., *Science* **234**, 1526 (1986)].
48. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 655 (1989).
49. H. M. Berman et al., *Biophys. J.* **63**, 751 (1992); A. A. Gorin, V. Zhurkin, W. K. Olson, *J. Biomol. Struct. Dyn.* **12**, a074 (1995).
50. The coordinates for the LEF-1-DNA complex of Love et al. (7) were not made available to us at the time of writing. The structure of both the protein and the DNA, however, are nearly identical to the structure of the SRY-DNA complex (6) at the site of intercalation.
51. N. Boutonnet, X. Hui, K. Zakrzewska, *Biopolymers* **33**, 479 (1993); E. Stofer and R. Lavery, *ibid.* **34**, 337 (1994).
52. M. Carson, *J. Appl. Crystallogr.* **24**, 958 (1991); *J. Mol. Graphics* **5**, 103 (1987).
53. A. J. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
54. The style of base representations is adapted from C. R. Calladine and H. R. Drew, *Understanding DNA: The Molecule and How It Works* (Academic Press, New York, 1992).
55. We wish to acknowledge W. A. Eaton, A. Szabo, H. A. Nash, and V. Zhurkin for critical reading of the manuscript; J. Aarons for help with the art work in Fig. 3; V. Zhurkin for assistance with the analysis of distorted DNA structures and many stimulating discussions; A. A. Gorin, V. Zhurkin, and W. K. Olson for the use of CompDNA; R. G. Brennan, W. I. Lipscomb, S. K. Burley, and X. Cheng for providing crystal structure coordinates; and M. Summers, R. J. Fisher, M. Brenowitz, P. Rice, S.-W. Yang, K. Mizuuchi, and H. A. Nash for communicating information before publication. Supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.).

AAAS–Newcomb Cleveland Prize

To Be Awarded for a Report, Research Article, or an Article Published in *Science*

The AAAS–Newcomb Cleveland Prize is awarded to the author of an outstanding paper published in *Science*. The value of the prize is \$5000; the winner also receives a bronze medal. The current competition period began with the 2 June 1995 issue and ends with the issue of 31 May 1996.

Reports, Research Articles, and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration for the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the competition period, readers are

invited to nominate papers appearing in the Reports, Research Articles, or Articles sections. Nominations must be typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to the AAAS–Newcomb Cleveland Prize, AAAS, Room 924, 1333 H Street, NW, Washington, DC 20005, and **must be received on or before 30 June 1996**. Final selection will rest with a panel of distinguished scientists appointed by the editor-in-chief of *Science*.

The award will be presented at the 1997 AAAS annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.