DNA recognition and bending

Rudolf K Allemann[1] and Martin Egli[2]

DNA-binding proteins recognize their DNA targets not only through the formation of specific contacts with the nucleotide bases but also through inherent properties of the DNA sequence, including increased bendability and rigidity. Consideration of the properties of both the protein and the DNA is required before the sequence specificity and the observed DNA bend in DNA–protein complexes can be understood.

Addresses: [1]Laboratory for Organic Chemistry, Department of Chemistry, ETH-Zurich, Universitätstrasse 16, CH-8092 Zurich, Switzerland. [2]Department of Molecular Pharmacology and Biological Chemistry and Drug Discovery Program, Northwestern University Medical School, Chicago, IL 60611-3008, USA.

Correspondence: Rudolf K Allemann and Martin Egli
E-mail: allemann@chem.ethz.ch and m-egli@nwu.edu

Chemistry & Biology September 1997, **4**:643–650
http://biomednet.com/elecref/1074552100400643

Introduction

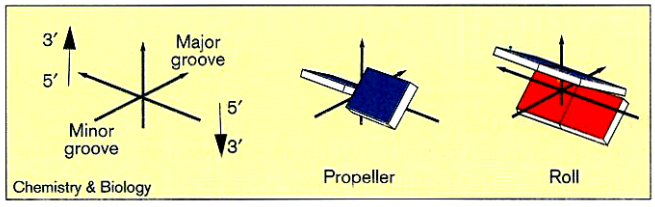
Any transcription factor must be able to recognize and bind to its target DNA efficiently and specifically. The precise recognition of a defined DNA sequence by a given transcription factor necessitates an optimal shape complementarity between the interacting species, whereby both the protein and the DNA can adapt their conformation to ensure an optimal fit. Many proteins recognize B-DNA through the formation of a number of hydrogen bonds and van der Waals interactions between amino acid sidechains and the functional groups of the bases. On the other hand, it has long been known that DNA can adopt a bent conformation when bound to a protein [1]. Both crystal-structure analyses of synthetic DNA oligonucleotides [2,3] and gel-electrophoresis experiments [4–7], however, have shown that certain DNA sequences can be bent even in the absence of proteins. Here, we discuss the effect of nucleotide sequence on DNA bending and how interacting proteins can influence the conformation of DNA so as to achieve an optimal fit with the protein.

Sequence-dependent DNA bending

The one strong determinant of intrinsic DNA bending is the nucleotide sequence. DNA with the general sequence RGCY (where R = A or G and Y = C or U) appears always to be bent at the central GC step because guanines and cytosines prefer to stack squarely on top of each other. For instance, in the crystal structure of the oligonucleotide d(CATGGCCATG), a bend angle of 23° was observed at the GC step in the DNA helix [8]. The three hydrogen bonds that hold a G–C base pair together prevent the formation of a significant propeller twist between the bases, which in turn prevents the formation of a roll between adjacent base pairs (Figure 1; [9,10] and references therein). The stacking can therefore only continue for a short distance before the winding of the helix requires the stacking to be diminished [3] and, consequently, the helix is bent at this position.

Another type of sequence-dependent bending occurs at the junctions between regions of G–C and A–T base pairs. DNA bending of 10–20° has been observed in the crystal structures of oligonucleotides that have an AT core ([3] and references therein). Although such sequences may be bent, they are not always so. The transition from G–C to A–T base pairs renders such a region of the DNA flexible and therefore able to bend. Such facultative bending is illustrated by the X-ray structure of the dodecamer d(CGCGAATTCGCG); it has a bend of 18° at the G–C/A–T junction at one end of the helix but is unbent at the equivalent junction at the other end of

Figure 1

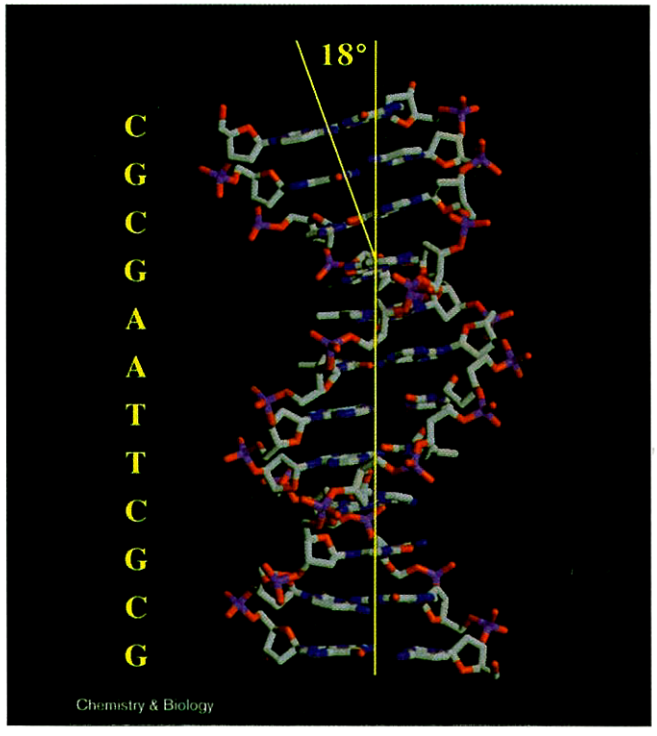


The propeller twist and the roll in a DNA helix. Propeller twisting is a property of a single base pair and rolling occurs between two adjacent base pairs. The roll is positive if the base pairs open up towards the minor-groove side (as shown). A positive propeller twist of a base pair corresponds to a clockwise rotation between bases.

the helix (Figure 2) [11,12]. The helical bend is generated by positive roll angles (Figure 1) of 8.8° and 5.3° at the CG and GA steps, respectively (all helical parameters were calculated using the program NEWHEL93 distributed by R.E. Dickerson and are based on the recent 1.4 Å resolution crystal structure data of the Dickerson–Drew DNA dodecamer; G. Hu, X. Shui and L.D. Williams, personal communication). The increased propeller twist in the AT region, which generates a reduced minor-groove width of approximately 9.5 Å (Figure 2), is possible because A–T base pairs are held together by only two hydrogen bonds rather than the three formed between guanine and cytosine. The propeller twist is large enough to allow the formation of interstrand bifurcated hydrogen bonds across the major groove between the 6-amino group of adenine and the O(4) of thymine of the neighbouring base pair in the AA step [13–17]. Although this interaction is not necessarily required for DNA bending [18], as is obvious from the bending observed in DNA with alternating adenines and thymines [19], it clearly stabilizes the high propeller twist in AT regions.

The exceptionally small minor-groove width in AT regions has also been observed in the crystal and nuclear magnetic resonance (NMR) structures of DNA complexed with various minor-groove-binding drugs [20–24]. In addition, the absence of a 2-amino group in adenine facilitates bending into the minor groove. Removal of the 2-amino group of guanine by replacing guanine with inosine (I) reduces the electrophoretic mobility of DNA sequences, which indicates that the intrinsic curvature of the DNA is increased. On the other hand, replacing adenine in natural DNA with 2,6-diaminopurine reduced the intrinsic DNA curvature [25]. The importance of the lack of a 2-amino group for bending AT regions into the minor groove has also been demonstrated by the X-ray structure analysis of the drug distamycin bound to the minor groove of an oligonucleotide with the sequence $d(IC)_4$ [26]. The presence of inosine, lacking a 2-amino group, rather than guanine, makes the I–C base pairs behave like A–T base pairs.

Figure 2



A facultative bend into the major groove at one end of the dodecamer duplex $[d(CGCGAATTCGCG)]_2$ [12]. Note the positive roll at the CG and GA steps and the negative propeller twisting in the central four base pairs. Overall and local helical axes are shown in yellow. Carbon atoms, grey; nitrogen atoms, dark blue; oxygen atoms, red; phosphorus atoms, purple.

The bendability of a TA step is inherently much greater than that of an AT or an AA step. A simple mechanical model to explain this observation has been provided by Finch and coworkers [27]. The proximity of the methyl groups of two successive thymines to each other and to the phosphate backbone makes the helix rather rigid (see Figure 3 in [28]). In an AT step, the stacking of the methyl group of the thymine with the adjacent adenine and the intervening sugar phosphate backbone again prevents bending by a roll mechanism. In a TA step, however, the methyl group projects into the major groove without any significant stacking interactions with either the adjacent adenine or the phosphate backbone, making this step more flexible than the AA and AT steps.

Despite these structural observations on the intrinsic ability of DNA to bend, protein–DNA complexes have often been viewed as a simple linear DNA duplex decorated with a protein. Often, the DNA in a DNA–protein complex is indeed straight. The early X-ray analyses of phage repressors [28,29] have revealed that these proteins achieve their high degree of DNA-binding specificity through extensive sets of contacts between amino acid

Table 1

DNA-binding proteins, their binding sites and the DNA bend angles in the DNA–protein complexes.

Protein	Recognition sequence	Bend angle (°)	Reference
CAP	CGAAAAGTGTGACAT-ATGTCACACTTTTCG	90	[32]
SRF	CTAATTAG	72	[34]
TBP	CTATAAAAG	100	[46–48]
MEF-2C	CTATAAATAG	~50	*
*Eco*RI	GAATTC	~30	[52]
M.*Eco*RI	GAATTC	~50	[56]
*Eco*RV	GATATC (cognate)	50	[53]
*Eco*RV	GAGCTC (non-cognate)	no bend	[53]
*Fok*I	GGATG	no bend	[55]
M.*Hha*I	GCGC	no bend	[57]
E47	CACCTG	no bend	[61]
Max	CACGTG	no bend	[62]
USF	CACGTG	no bend	[63]
PurR	ACGAAAACGTTTTCGT	45	[65]
SRY	GCACAAAC	65	[66]
ETS1	TCGAGGCCGGAAGTTCGA	60	[67]
IHF	AAAAAAGCATTGCTT-ATCAATTTGTTGCA	160	[69]

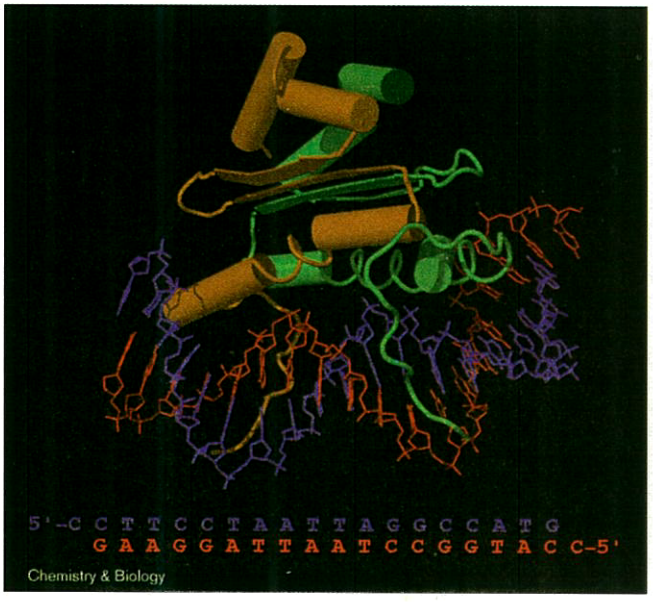
*D. Meierhans, M. Sieber and R.K A., unpublished observations. SRF, serum response factor; TBP, TATA-binding protein; *Eco*RI, M.*Eco*RI, *Eco*RV, *Fok*I, restriction endonucleases; MEF-2C, myocyte enhancer factor-2C; E47, basic helix–loop–helix transcription factor; Max, USF, basic helix–loop–helix-leucine zipper proteins; PurR, a purine repressor; SRY, a sex-determining protein; ETS1, a human oncogene product; IHF, integration host factor.

residues and the operator bases. For instance, the complex of the bacteriophage λ repressor with its DNA-binding site has 16 hydrogen bonds between sidechains and the nucleotide bases in addition to numerous van der Waals interactions [30]. The amount of DNA bending in phage repressor–DNA complexes is small.

DNA bending in DNA–protein complexes

The first examples of extensively bent DNA structures were found in the complexes of DNA and the dimeric proteins bacteriophage λ Cro [31] and catabolite activator protein (CAP) [32]. Each CAP monomer recognizes one half site of its DNA target through the introduction of an α helix into the major groove, thereby compressing the minor groove between the two insertion sites and bending the DNA by approximately 90° (Table 1) [32]. Cro also recognises its DNA target sequence TATCACCGCGGGT-GATA through a helix–turn–helix motif. Although the first and last six base pairs in the operator essentially have a B-type conformation, the central region of the DNA is bent

Figure 3



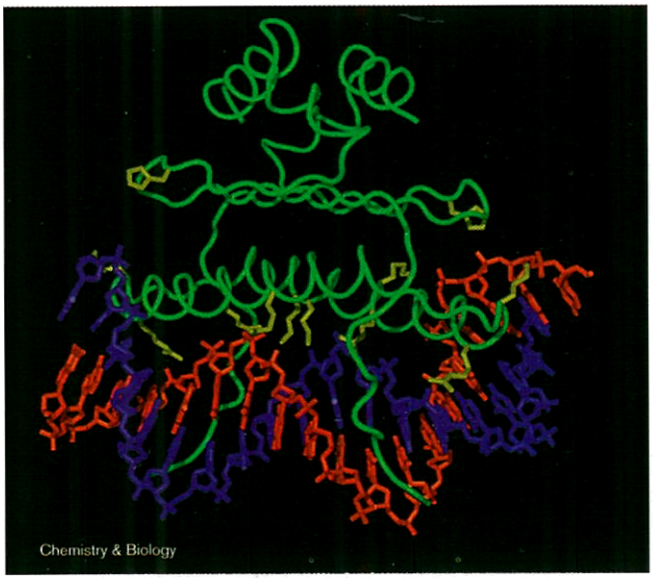
The global conformation of the SRF–DNA complex [34]. The protein dimer binds from the minor-groove side and bends the DNA into the minor groove. The protein monomers are colored in orange and green, α helices are represented as cylinders and β strands are shown as arrows. The strands of the DNA duplex are colored in red and blue, and their sequences are indicated below.

by approximately 40° and the five G–C base pairs in the middle of this region are significantly overwound [31]. These dramatic bending effects would not have been predicted by a simple analysis of the sequences of the binding sites of CAP and Cro.

Conversely, the recognition sequences of serum response factor (SRF), TATA-binding protein (TBP), the restriction endonucleases *Eco*RI and *Eco*RV, and *Eco*RI DNA methyltransferase (M.*Eco*RI) all have AT-rich cores (Table 1), characterized by an increased bendability even in the absence of a protein. The crystal structure analyses of DNA complexed to these proteins revealed that they exploit the inherent bendability of their DNA targets, and the DNA is significantly bent in all these complexes (Table 1).

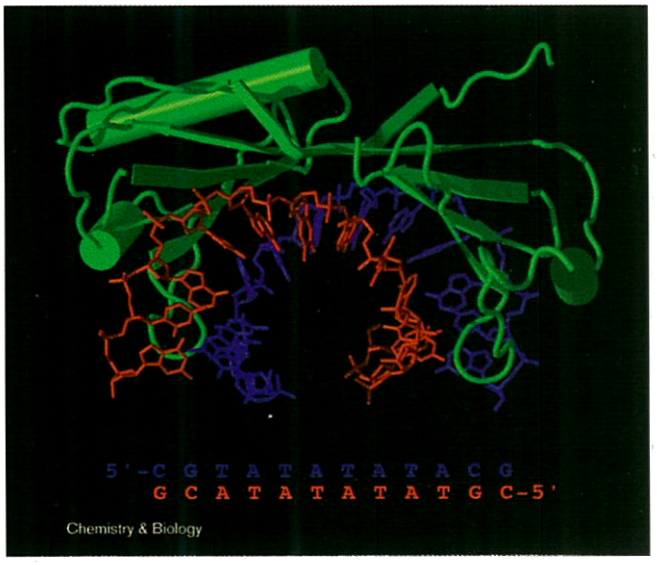
SRF is a ubiquitous nuclear protein that is important for cellular proliferation and differentiation [33]. It binds to DNA as a homodimer through a MADS domain and depends on the presence of the DNA sequence $CC(A/T)_6GG$ for activity (Figure 3). A determination of the crystal structure of the SRF–DNA complex showed that the DNA was bent by 72° as a result of high positive roll angles at the G–C/T–A junction [34]. The major groove is opened up to 19.5 Å in the AT region, whereas the minor groove is compressed to approximately 8 Å through significantly increased negative propeller twists.

Figure 4



Interactions between the positively charged amino acid sidechains (yellow) and the DNA phosphates in the SRF–DNA complex; the remainder of the protein is shown in green. Note that these contacts occur on only one side of the duplex, causing the DNA (red and blue) to bend towards the protein. The histidine and lysine residues attached to either end of the duplex act as handles to bend the DNA further (the DNA illustrated here is too short to allow formation of the His–phosphate contact on the left-hand side).

Figure 5



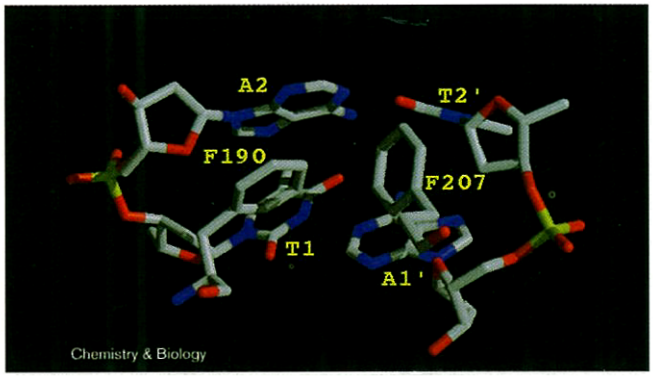
The global conformation of the TBP–DNA complex [47]. The protein binds the TATA box from the minor-groove side and bends the DNA into the major groove. The α helices and β strands of the protein (green) are represented as cylinders and arrows, respectively. The sequence of the DNA duplex (red and blue) is shown.

Negative roll angles were observed between the central four base pairs of the SRF-binding site, which reduce the unfavourable interaction between the stacked exocyclic amino groups of adjacent adenines [34]. As expected, binding-site selection experiments based on the polymerase chain reaction (PCR) revealed a strong conservation of the flanking GG dinucleotide and a discrimination against guanine and cytosine in the core of the SRF-binding site [35]. Mutations in the core of the binding site are tolerated as long as they do not prevent the formation of the high propeller twists required in this region. Overall, the conformation of the DNA in the complex is similar to that expected for the unbound DNA and SRF appears simply to potentiate the high intrinsic bendability of its DNA target site [34]. Several positively charged amino acid residues of SRF bind to the phosphate groups on only one side of the SRF-binding site (Figure 4). These interactions lead to an unbalanced Coulombic repulsion between the phosphate groups, causing the DNA to bend towards SRF [36,37]. The three positively charged amino acids that interact with the distal ends of the DNA pull these regions of the DNA up towards SRF.

Myocyte enhancer factor-2C (MEF-2C) interacts with DNA in a similar fashion to SRF [38]. Target sites for MEF-2C have been identified in the promoters and enhancers of many muscle-specific genes [39]; the consensus DNA-binding site was determined as $CTA(A/T)_4TAG$ [38]. A detailed analysis of the affinity of MEF-2C for various DNA sequences revealed that mutations within the central four base pairs are tolerated as long as adenine is replaced with thymine or *vice versa* (D. Meierhans, M. Sieber and R.K.A., unpublished observations). The replacement of the central bases with guanine and cytosine, however, diminished the binding affinity significantly. Circular dichroism (CD) spectroscopy and bending analysis by circular permutation assays revealed that MEF-2C indeed potentiates the natural tendency of its DNA target to adopt a bent conformation (Table 1; D. Meierhans, M. Sieber and R.K.A., unpublished observations).

The TATA-binding protein (TBP) uses another mechanism to increase the bending of DNA that already has a natural propensity to bend [40]. TBP is a phylogenetically highly conserved, central transcription factor required for transcriptional initiation by all three eukaryotic RNA polymerases [41–44]. The first step in gene transcription by RNA polymerase II is the binding of TBP to a TATA box, a highly conserved AT-rich promoter element. The consensus TATA-box sequence is TATA(A/T)A(A/T) [45]. The crystal structures of a TATA box complexed with human, yeast and *Arabidopsis* TBPs have recently been solved (Figure 5) [46–48]. Strong kinks of approximately 45° are introduced into the DNA at either end of the TATA box through the intercalation of two phenylalanine rings between two adjacent base pairs (Figure 6),

Figure 6



One of the two sharp kinks which bend the DNA towards the major groove in the TBP–DNA complex. The kinks are introduced by TBP through the insertion of two phenylalanines (F) between adjacent T–A pairs from the minor-groove side [48]. Note the massive positive roll of more than 40° between the base pairs. Carbon atoms, grey; nitrogen atoms, blue; oxygen atoms, red; phosphorus atoms, yellow.

Table 2

A comparison of the bend angles in various TATA-box–TBP complexes, and the relative activity of transcriptional activation in HeLa and yeast cells.

Sequence	Bend angle*	Relative *in vitro* activity†	
		HeLa	Yeast
TATAAA (A)	93	100	100
TATATAA	106	172	107
TTTAAA (A)	87	54	62
TATTAA (A)	80	19	25
CATAAA (A)	63	26	30
TAAAAA (A)	< 34	< 1	< 1
TACAAA	nd	2	2
TAGAAA	nd	2	1
TATGAA	nd	< 1	< 1

*See [50]. †See [51]. nd, not determined.

thereby bending the DNA into the major groove (Figure 5). The bends are produced by rolling these adjacent base pairs around their long axes, resulting in a compression of the major groove and a widening of the minor groove (Figure 5).

The activity of TBP in yeast and HeLa (a human epithelial cell line) cells was measured with different TATA boxes. It could be correlated with the bendability of the TATA box and with the bend angle observed in the TBP–DNA complex (Table 2) [49,50]. Transcription from the sequence TATAAA was used as a standard. The bend angle in the complex with TBP was independently determined to be approximately 90° by X-ray crystallography [48] and by solution methods [49]. Increasing the ability of the DNA to bend through the introduction of an additional TA step (TATATAA) also increased both the transcriptional activity and the observed bend angle in the complex (Table 2). The simple inversion of an A–T base pair generally affects both the transcriptional activity and the bend angle only slightly. The inversion of the A–T base pair at position 3, however, creates a pure A tract (TAAAAA), which is typically rigid (see above), and both the bend angle and the transcriptional activity are dramatically decreased. The only other mutation that creates a similar reduction in activity is the introduction of a G–C pair in the center of the TATA box.

The most remarkable property of type II restriction enzymes is that they cut their target DNA with high sequence specificity [51]. The restriction enzymes that recognize AT-rich DNA sequences provide another striking example of a correlation of the bendability of the free DNA with the bending observed in the complex. *Eco*RI and *Eco*RV, which recognize GAATTC and GATATC, respectively, bend the DNA by 30° and 50° through an unstacking of the central base step (Table 1; [52] and references therein; [53]). In the crystal structure of the complex between *Eco*RV and the non-cognate DNA sequence GAGCTC, however, the DNA was essentially unbent (Table 1), presumably because a higher expenditure of free energy would be needed to distort the non-cognate DNA [53]. Replacing the second adenine in the *Eco*RI recognition sequence with purine (GAPTTC) led to an increase in binding affinity, probably because the formation of a high propeller twist in the P–T base pair is easier than in the A–T base pair [54]. The determination of the crystal structure of DNA complexed with the restriction endonuclease *Fok*I, which recognises the asymmetric DNA sequence GGATG, revealed that the DNA maintains an unbent B-type conformation [55].

Fluorescence spectroscopy and crystal-structure analysis, respectively, showed that both M.*Eco*RI and *Hha*I DNA-methyltransferase (M.*Hha*I) stabilize an extrahelical base [56,57]. For DNA recognition, however, M.EcoRI takes advantage of the intrinsically high flexibility of its recognition sequence [58], whereas the conformation of the M.*Hha*I-binding site remains unbent upon protein binding, as might have been predicted from its recognition sequence (GCGC; Table 1).

In all the cases described above, the protein takes advantage of the bendable nature of AT sequences. To the best of our knowledge, the only case in which such an inherently bendable DNA sequence adopts a straight conformation in a protein complex was found in the cocrystal structure of the architectural transcription factor HMG-I(Y) (high-mobility group protein) with an oligonucleotide containing

the PRDII (positive regulatory domain 2 of the human interferon-β enhancer) site (GGGAAATTCCTC) of the interferon-β promoter [59]. The interferon-β enhancer has a small intrinsic bend of ~20° towards the minor groove [60]. On the other hand, the PRDII site in the complex with HMG-I(Y) has an essentially straight B-type conformation [59]. It appears that the principal role of HMG-I(Y) is to reverse or prevent the distortion of a DNA site that is inherently bendable. HMG-I(Y) achieves this through the formation of numerous contacts with the phosphate backbone and the nucleotide bases in the minor groove [59].

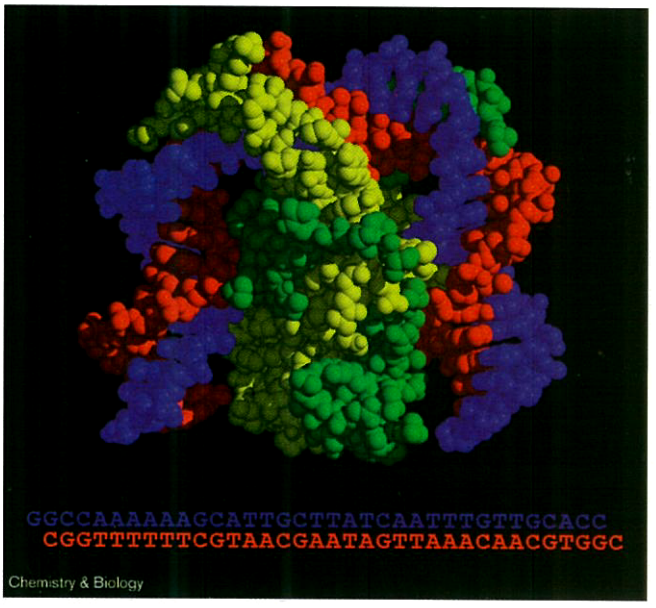
When a protein binds to a more rigid DNA sequence it often does not bend the DNA. M.*Hha*I [57], E47 (basic helix–loop–helix transcription factor) [61], and the two basic helix–loop–helix-leucine zipper proteins Max [62] and USF [63] bind to those DNA sequences that are most likely to adopt a B-type conformation in the uncomplexed form, and indeed the DNA remains unbent even when bound to a protein (Table 1).

Proteins can also induce the formation of bends in DNA sequences that are not inherently bendable but this is energetically costly. The purine repressor, PurR, which is involved in the biosynthesis of purines and pyrimidines, binds to runs of four consecutive adenines and thymines interrupted by the dinucleotide CG (Table 1) [64]. Like TBP, PurR bends the DNA into the major groove through the insertion of a hinge into the minor groove and the intercalation of the sidechains of two leucine residues into the CG step [65]. In addition, two helix–turn–helix motives bind into the major groove on either side of the CG step, thereby providing significant leverage for the bending motion, resulting in a 45° bend of the helical axis [65]. The AAAA regions that flank the central CG dinucleotide are straight, as shown by the high propeller twists and the small roll angles, which is in good agreement with the properties of A tracts (see above). The protein takes advantage of the typical characteristics of A tracts such as the straight course of the DNA and the narrow minor groove.

The sex-determining protein SRY [66] and the human oncogene product ETS1 [67], which bend the helical axis by 65° and 60°, respectively (Table 1), are other proteins that bind to DNA sequences that are not inherently bendable but nevertheless adopt bent conformations in the protein–DNA complex. In both cases, intercalation of amino acid residues between base pairs of the DNA and extensive other contacts are needed to bend the DNA [40]. Sharp DNA bends, irrespective of the exact base sequence and the inherent ability of DNA to bend, are also found in the nucleosome core particle generated by extensive contacts between histones and DNA [68].

A further example of energetically demanding protein-induced DNA bending is provided by the structure of

Figure 7



A van der Waals' representation of the IHF–DNA complex [70]. The monomeric subunits of the protein are shown in yellow and green. The strands of the DNA duplex are red and blue and its sequence is depicted below. The drawing illustrates both the massive DNA bend induced by the protein and the large contact surface of ~4600 Å^2 between the protein and the DNA. Note the straight A tract on the left-hand side with one IHF momomer (green) contacting its narrow minor groove.

integration host factor (IHF) bound to the H′ site of phage λ (GCCAAAAAAGCATTGCTTATCAATTTGTTGCACC) (Figure 7) [69,70]. The 34 base pair DNA is literally wrapped around the protein and bent by more than 160°, thereby reversing the direction of the DNA within a very short distance. For bending, IHF relies on the intercalation of proline residues and extensive interactions with the DNA, creating a buried protein–DNA interface of ~4600 Å^2 (Figure 7). Such an extreme bend, which allows proteins that bind upstream and downstream of the IHF-binding site to interact directly with one another, provides the structural basis for the function of IHF in processes like λ integration, replication and site-specific recombination (for reviews, see [71,72]). The interaction of IHF with the A tract at one end of the DNA duplex provides another striking example of recognition through structure rather than through base-specific contacts. The structural features of the A tract are exploited by IHF; the narrow minor groove creates an optimal shape complementarity between the protein and the DNA. In fact, the A tract in the complex with IHF can be superimposed on that of the dodecamer $[d(CGCAAAAAAGCG)]_2$ [13] with a root mean square deviation of 0.6 Å (Figure 7) [69].

Conclusions

In summary, there is ample evidence that proteins recognize and bind to their specific DNA targets not only

through sequence-specific contacts between the amino acid residues and the bases of the DNA, but also through the recognition of intrinsic properties of the DNA-binding site such as the tendency to bend. The conservation of flexibility or rigidity is therefore at least as important for recognition as the strict conservation of the exact DNA sequence. The bending properties and the specificity of DNA–protein complexes are best understood by studying the properties of both the protein and the DNA.

Acknowledgements

We thank G. Marius Clore for a preprint of the manuscript describing the structure of the DNA complex of HMG-I/Y, Martin Sieber for preparing the figures, Valentina Tereshko for calculating the helical parameters used in this review, Neda Mashhoon for drawing our attention to DNA bending in the M.*Eco*RI complex, and Wayne F. Anderson and Lesley A. Tannahill for critical reading of the manuscript and for helpful comments.

References

1. Klug, A., Jack, A., Viswamitra, M.A. Kennard, O., Shakked, Z. & Steitz, T.A. (1979). A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the *lac*-repressor protein. *J. Mol. Biol.* **131**, 669-680.
2. Dickerson, R.E., Goodsell, D. & Neidle, S. (1994). "...the tyranny of the lattice ...". *Proc. Natl Acad. Sci. USA* **91**, 3579-3583.
3. Dickerson, R.E., Goodsell, D. & Kopka, M.L. (1996). MPD and DNA bending in crystals and in solution. *J. Mol. Biol.* **256**, 108-125.
4. Kahn, J.D., Yun, E. & Crothers, D.M. (1994). Detection of localised DNA flexibility. *Nature* **368**, 163-166.
5. Koo, H.-S., Wu, H.-M. & Crothers, D.M. (1986). DNA bending at adenine·thymine tracts. *Nature* **320**, 501-506.
6. Wu, H.-M. & Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-513.
7. Zinkel, S.S., & Crothers, D.M. (1987). DNA bend direction by phase sensitive detection. *Nature* **328**, 178-181.
8. Goodsell, D.M., Kopka, M.L., Cascio, D. & Dickerson, R.E. (1993). Crystal structure of CATGGCATG and its implications for A-tract bending models. *Proc. Natl Acad. Sci. USA* **90**, 2930-2934.
9. Dickerson, R.E. (1983). The DNA helix and how it is read. *Sci. Am.* **249**, 94-102.
10. Calladine, C.R. & Drew, H.R. (1997). *Understanding DNA. The Molecule and How it Works.* Academic Press Inc., San Diego, CA.
11. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755-758.
12. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA* **78**, 2179-2183.
13. Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987). The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature* **330**, 221-226.
14. Coll, M., Frederick, C.A., Wang, A., H.-J. & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl Acad. Sci. USA* **84**, 8385-8389.
15. DiGabriele, A.D., Sanderson, M.R. & Steitz, T.A. (1989). Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl Acad. Sci. USA* **86**, 11816-11819.
16. DiGabriele, M.R. & Steitz, T.A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024-1039.
17. Shatzky-Schwartz, M., *et al.*, & Shakked, Z. (1997). X-ray and solution studies of DNA oligomers and implications for the structural basis of a-tract dependent curvature. *J. Mol. Biol.* **267**, 595-623.
18. Diekmann, S., Mazzarelli, J.M., McLaughlin, L.W., von Kitzing, E. & Travers, A.A. (1992). DNA curvature does not require bifurcated hydrogen bonds or pyrimidine methyl groups. *J. Mol. Biol.* **225**, 729-738.
19. Yoon, C., Privé, G.G., Goodsell, D.S. & Dickerson, R.E. (1988). Structure of an alternating B-DNA helix and its relationship to A-tract DNA. *Proc. Natl Acad. Sci. USA* **85**, 6332-6336.
20. Kopka, M.L. & Larsen, T.A. (1992). Netropsin and the lexitropins. The search for sequence-specific minor-groove-binding ligands. In *Nucleic Acid Targeted Drug Design.* (Probst, C.L. & Perun, T.J., eds) pp. 303–374, Marcel Dekker, Inc., New York.
21. Brown, D.G., Sanderson, M.R., Garman, E. & Neidle, S. (1992). Crystal structure of a berenil-d(CGCAAATTTGCG) complex: an example of drug-DNA recognition based on sequence-dependent structural features. *J. Mol. Biol.* **226**, 481-490.
22. Goodsell, D.S., Kopka, M.L. & Dickerson, R.E. (1995). Refinement of netropsin bound to DNA: bias and feedback in electron density map interpretation. *Biochemistry* **34**, 4983-4993.
23. Nunn, C.M., Garman, E. & Neidle, S. (1997). Crystal structure of the DNA decamer d(CGCAATTGCG) complexed with the minor groove binding drug netropsin. *Biochemistry* **36**, 4792-4799.
24. Neidle, S. (1997). Crystallographic insights into DNA minor groove recognition by drugs. *Biopolymers* **44**, 105-121.
25. Bailly, C., Waring, M.J. & Travers, A.A. (1995). Effects of base substitutions on the binding of a DNA-bending protein. *J. Mol. Biol.* **253**, 1-7.
26. Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingam, M. (1994). Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nat. Struct. Biol.* **1**, 169-175.
27. Suzuki, M., Yagi, N. & Finch, J.T. (1996). Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* **379**, 148-152.
28. Aggarwal, A.K., Rogers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of Phage 434: a view at high resolution. *Science* **242**, 899-907.
29. Jordan, S.R. & Pabo, C.O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* **242**, 893-899.
30. Beamer, L.J. & Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* **227**, 177-196.
31. Brennan, R.G., Roderick, S.L., Takeda, Y. & Matthews, B.W. (1990). Protein-DNA conformational changes in the crystal structure of a λ Cro-operator complex. *Proc. Natl Acad. Sci. USA* **87**, 8165-8169.
32. Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253**, 1001-1007.
33. Norman, C., Runswick, M., Pollock, R. & Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **55**, 989-1003.
34. Pellegrini, L., Tan, S. & Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. *Nature* **376**, 490-498.
35. Pollock, R. & Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18**, 6197-6204.
36. Rich, A. (1978). Localised positive charges can bend double helical nucleic acid. *Fed. Eur. Biochem. Soc.* **51**, 71-81.
37. Strauss, J.K. & Maher III, L.J. (1994). DNA bending by asymmetric phosphate neutralisation. *Science* **266**, 1829-1834.
38. Shore, P. & Sharrock, A.D. (1995). The MADS-box family of transcription factors. *Eur. J. Biochem.* **229**, 1-13.
39. Gosset, L.A., Kelvin, D.J., Sternberg, E.A. & Olson, E.N. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol. Cell. Biol.* **9**, 5022-5033.
40. Werner, M.H., Gronenborn, A.M. & Clore, G.M. (1996). Intercalation, DNA kinking, and the control of transcription. *Science* **271**, 778-784.
41. Sharp, P.A. (1992). TATA-binding protein is a classless factor. *Cell* **68**, 819-821.
42. Rigby, P.W.J. (1993). Three in one and one in three: it all depends on TBP. *Cell* **72**, 7-10.
43. Tansey, W.P. & Herr, W. (1997). TAFs: guilt by association? *Cell* **88**, 729-732.
44. Marsh, T.L., Reich, C.I., Whitelock, R.B. & Olsen, G.J. (1992). Transcription factor IID in the archaea: sequences in the thermococcus celer genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl Acad. Sci. USA* **91**, 4180-4184.
45. Bucher, P. (1990). Weight matrix description of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563-578.
46. Juo, Z.S., Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. & Dickerson, R.E. (1996). How proteins recognize the TATA box. *J. Mol. Biol.* **261**, 239-254.

47. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-529.
48. Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527.
49. Starr, D.B., Hobbs, B.C. & Hawley, D.K. (1995). DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* **250**, 434-446.
50. Wobbe, C.R. & Struhl, K. (1990). Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol. Cell. Biol.* **10**, 3859-3867.
51. Roberts, R.J. & Halford, S.S. (1993). Type II restriction endonucleases. In *Nucleases*. (Linn, S.M., Lloyd, R.S. & Roberts, R.J., eds), pp. 35-88, Cold Spring Harbor Laboratory Press, New York.
52. Kim, Y., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. (1990). Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307-1309.
53. Winkler, F.K., *et al.*, & Wilson, K.S. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781-1795.
54. Lesser, D.R., Kurpiewski, M.R., Waters, T., Connolly, B.A. & Jen Jacobson, L. (1993). Facilitated distortion of the DNA site enhances EcoRI endonuclease-DNA recognition. *Proc. Natl Acad. Sci. USA* **90**, 7548-7552.
55. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100.
56. Allan, B.W. & Reich, N.O. (1996). Targeted base stacking disruption by the EcoRI DNA methyltransferase. *Biochemistry* **35**, 14757-14762.
57. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.
58. Garcia, R.A., Bustmante, C.J. & Reich, N.O. (1996). Sequence-specific recognition by cytosine C^5 and adenine N^6 DNA methyltransferases requires different deformations of DNA. *Proc. Natl Acad. Sci. USA* **93**, 7618-7622.
59. Huth, J.R., *et al.*, & Clore, G.M. (1997). The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657-665.
60. Falvo, J.V., Thanos, D. & Maniatis, T. (1995). Reversal of intrinsic DNA bends in the IFNβ gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83**, 1101-1111.
61. Ellenberger,T., Fass, D., Arnaud, M. & Harrison, S.C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **8**, 970-980.
62. Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B. & Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**, 38-45.
63. Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G. & Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. *EMBO J.* **13**, 180-189.
64. Zalkin, H. & Dixon, J.E. (1992). De novo purine nucleotide biosynthesis. *Prog. Nucleic Acid. Res. Mol. Biol.* **42**, 259-287.
65. Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science* **266**, 763-770.
66. Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705-714.
67. Werner, M.H., *et al.*, & Gronenborn, A.M. (1995). The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain interaction. *Cell* **83**, 761-771.
68. Richmond, T.J., Finch, J.T., Rhodes, D. & Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532-537.
69. Rice, P.A., Yang, S.-W., Mizuuchi, K. & Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-Turn. *Cell* **87**, 1295-1306.
70. Yang, S.-W. & Nash, H.A. (1995). Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. *EMBO J.* **14**, 6292-6300.
71. Nash, H.A. (1996). The HU and IHF proteins: accessory factors for complex protein-DNA assemblies. In *Regulation of Gene Expression in* Escherichia coli. (Lin, E.C.C. & Lynch, A.S., eds), pp. 149–179, R.G. Landes Company, Austin, TX.
72. Goosen, N. & van de Putte, P. (1995). The regulation of transcription initiation by integration host factor. *Mol. Microbiol.* **16**, 1-7.