Annu. Rev. Biophys. Biomol. Struct. 1998. 27:105–31

MINOR GROOVE-BINDING ARCHITECTURAL PROTEINS: Structure, Function, and DNA Recognition[1]

Carole A. Bewley, Angela M. Gronenborn, G. Marius Clore
Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-520; e-mail: clore@vger.niddk.nih.gov

KEY WORDS: TATA box binding protein, HMG-box proteins, integration host factor, HMG I(Y), DNA bending

ABSTRACT

To date, high-resolution structures have been solved for five different architectural proteins complexed to their DNA target sites. These include TATA-box–binding protein, integration host factor (IHF), high mobility group I(Y)[HMG I(Y)], and the HMG-box–containing proteins SRY and LEF-1. Each of these proteins interacts with DNA exclusively through minor groove contacts and alters DNA conformation. This paper reviews the structural features of these complexes and the roles they play in facilitating assembly of higher-order protein–DNA complexes and discusses elements that contribute to sequence-specific recognition and conformational changes.

CONTENTS

[1]The US government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

INTRODUCTION

A recurring theme of events that control transcription, recombination, and replication is the involvement of multiprotein-DNA complexes. The intricate architecture of these complexes relies not only on sequence-specific protein-DNA interactions, but also on the interaction of multiple proteins whose DNA binding sites may be quite distant from one another. Thus, the formation of such higher-order structures usually requires that the conformation of the DNA template be bent or distorted in order to bring the requisite proteins into close proximity. In large part this task is accomplished by architectural proteins that typically lack the potential to activate transcription or carry out recombination on their own, but instead induce conformational changes in the DNA, thereby facilitating the assembly and enhancing the overall stability and activity of the multiprotein-DNA complexes.

Many DNA-binding proteins possessing disparate functions have been shown to bend their DNA recognition sites (28, 44, 45). A distinction can be made, however, between those proteins whose primary, if not sole, function is to bend DNA, and those proteins that happen to bend DNA in the process of carrying out their distinct primary functions. The first group we refer to as architectural proteins in the strictest sense, while the second group could include any other relevant DNA-binding proteins such as transcriptional activators or repressors. High resolution three-dimensional structures for five different architectural proteins complexed to their cognate DNA sites have been solved to date and include (*a*) TATA-binding protein (TBP) (23–26, 34), (*b*) the male sex determining factor SRY (59), (*c*) lymphoid enhancer-binding factor 1 (LEF-1) (30), (*d*) integration host factor (IHF) (39), and (*e*) high mobility group I(Y) [HMG I(Y)] (21). All five of these architectural proteins interact exclusively with the minor groove of DNA.

Although the deficiency of chemical features presented by the minor groove is typically considered to be insufficient for specific recognition, all of these minor groove-binding proteins bind with high affinity and varying degrees of sequence specificity. Moreover, they exhibit very different global folds (Figure 1) and use different strategies, or combinations of strategies, for recognition and binding. Accordingly, these proteins also remodel the DNA conformation in distinct

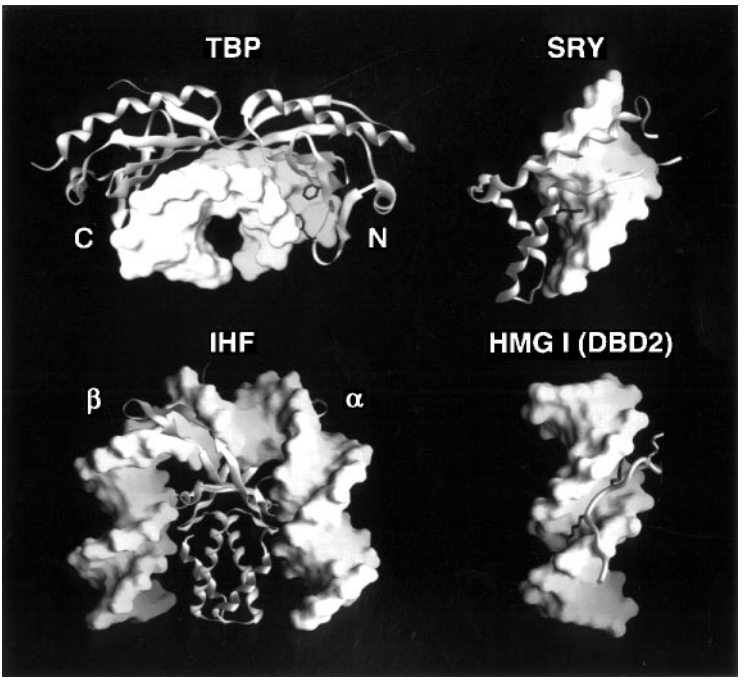

Figure 1 Three-dimensional structures of the minor-groove binding architectural proteins TBP, SRY, IHF, and HMG I in complex with their DNA targets. The DNA is represented as a surface and the protein as a ribbon. Where visible, sidechains key to modulating the DNA conformation are shown as *black rods*. These include Phe99 and Phe116 in the N-terminal domain of yTBP, Ile13 in SRY, and Arg10 and Arg12 of DNA-binding domain 2 of HMG I. The C- and N-terminal subunits of TBP and the α and β arms of IHF are labelled. Figures were produced with GRASP (32).

ways. The structures and functions of these minor groove binding architectural factors are reviewed below, and structural and chemical elements contributing to sequence-specific minor groove recognition are discussed.

STRUCTURE AND FUNCTION OF MINOR GROOVE-BINDING ARCHITECTURAL FACTORS

TATA-Box Binding Protein TBP

In eukaryotes, initiation of transcription of messenger, ribosomal, small nuclear, and transfer RNAs by any of the three RNA polymerases requires the TATA-box binding-protein (TBP) (19). In the case of class II nuclear genes in higher organisms, activation of transcription depends on the formation of a

functional preinitiation complex (PIC), comprising RNA polymerase II (polII) and a host of other general initiation factors (such as TFIIA, -B, -D, -E, -F, -I, -J), about the TATA element of core promoter sequences (3, 41, 43). TBP mediates PIC assembly because recognition of the TATA element by TBP is considered to be the nucleating event in complex formation. On its own, TBP cannot substitute for the accurate control of transcription provided by TFIID and larger transcription initiation complexes (20, 37), but it does recognize the TATA element, thereby initiating basal transcription (2).

The first view of a minor groove-binding architectural protein at work came in 1993 when two separate crystal structures of TBP complexed to TATA-box–containing oligonucleotides were solved by the groups of Sigler and Burley (25, 26): Sigler and coworkers solved the structure of yeast TBP bound to a 29-nucleotide DNA hairpin whose stem contains a 12-bp helix that encompasses the TATA box sequence 5′-TATATAAA-3′ from the yeast CYC1(-52) promoter, while Burley and coworkers solved the structure of TBP 2 from *Arabidopsis thaliana* bound to the TATA element of the adenovirus major late promoter (AdMLP) which contains the central binding sequence 5′-TATAAAAG-3′ (Table 1). Despite the differences in the protein sequences and their cognate

Table 1 Protein sources and DNA sequences used for structural studies of minor groove-binding architectural proteins

Protein source	DNA[a]	Reference
TBP *Arabidopsis thaliana*	5′GC**TATA**AAAGGGCA3′ 3′CG**ATAT**TTTCCCGT5′	24, 25
TBP yeast	5′G**TATA**TAAA ACGG$^{G}_{T}$ 3′C**ATAT**ATTT TGCG$_{G}$	26
TBP human	5′CTGC**TATA**AAAGGCTG3′ 3′GACG**ATAT**TTTCCGAG5′	34
TBP human	5′CG**TATA**TATACG3′ 3′GC**ATAT**ATATGC5′	23
SRY human	5′**GCACAAAC**3′ 3′**CGTGTTTG**5′	59
LEF-1 human	5′CAC**CCTTTGAA**GCTC3′ 3′GTG**GGAAACTT**CGAG5′	30
IHF *Escherichia coli*	5′GGCCAAAAAAGCATTGCT**TATCAA**TTTG**TTG**CACC3′ 3′CCGGTTTTTTCGTAACGA**ATAGTT**AAAC**AAC**GTGG5′	39
HMG I(Y) human	5′GGGAAATTCCTC3′ 3′CCCTTTAAGGAG5′	21

[a]Consensus sequences are shown in bold; required sequences such as A-tracts or AT-rich DNA are underlined. The sequences used in the TBP structures are aligned with respect to the 8 bp TATA elements.

DNA-binding sites, the primary structural features of these two protein-DNA complexes are nearly identical. Protein-DNA interactions for both complexes are summarized in Figure 2*a*, and numbering for the yeast TBP-TATA box complex is used in the text below unless otherwise specified.

TBP is composed of two separate domains, an amino-terminal domain that varies considerably in both length and sequence, and a 180-amino acid carboxy-terminal domain that is highly conserved. It has been shown that the amino-terminal domain is unnecessary for the initiation of transcription, and accordingly the carboxy-terminal domain has been shown to be sufficient for high affinity binding to TATA-box elements, as well as for initiating basal transcription (7). The carboxy-terminal domain of TBP is composed of two subunits, α and β, which contain 89 and 90 amino acids, respectively, and flank a highly basic region referred to as the basic repeat. Although the two subunits of the carboxy-terminal region display only 30% sequence identity, they are nearly identical structurally, and thus give rise to the strong intramolecular pseudo-symmetry of TBP (Figures 1, 3*b*); both crystal structures were solved using constructs of this carboxy-terminal domain bound to the TATA-box-containing promoter sequence. TBP is best described as having a saddle-like structure where each subunit of TBP consists of a five-stranded, curved antiparallel β-sheet and two α-helices, one long and one short. The central eight strands of the curved β-sheet form the concave underside of the saddle which interacts extensively with the distorted minor groove of the TATA-element, and the β-turns between the outermost antiparallel β-strands appear as stirrups of the saddle. The longer α-helix is located above and approximately orthogonal to the β-sheet, and together with the shorter helix provides a scaffold for the back, or convex side, of the β-sheet. As for the oligonucleotide, the central eight base-pairs that contain the TATA element are underwound and bent by $\sim 80^\circ$ toward the major groove, presenting a wide concave surface with complementary curvature to the underside of the protein (Figures 1, 3*b*). To either side of the TATA element, the DNA returns to B-form as it leaves the region of TBP binding.

Typically, transcription factors recognize and bind to DNA through specific contacts between protein sidechains and a variety of atoms or functional groups located in the major groove of DNA (36, 49). Although nonpolar interactions involving the methyls and olefinic protons of the pyrimidine bases are seen, the recognition interface is usually dominated by hydrogen bonds and electrostatic contacts between polar side chains of the protein and nitrogen and/or oxygen atoms of the DNA. In contrast the concave underside of TBP interacts with the minor groove of the TATA element, and somewhat surprisingly, the protein-DNA interface is dominated by hydrophobic or van der Waals contacts. Illustrative of this hydrophobic predominance, 70% of the 3000 Å^2 of buried surface area resulting from complex formation is hydrophobic. In addition, 12

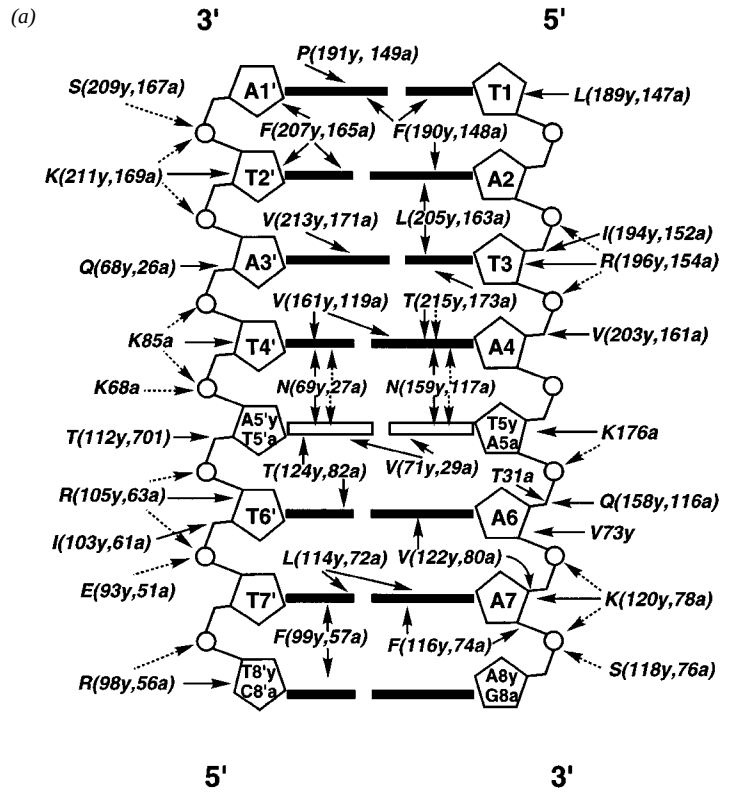

Figure 2 Schematic representations of protein-DNA contacts for each of the five minor groove-binding architectural proteins. The numbering for the DNA bases and the protein residues is the same as in the original papers, and the DNA is viewed looking into the minor groove. Deoxyribose rings are represented as *pentagons*, phosphates as *hollow circles*, and purine (A,G) and pyrimidine (C,T) bases as *long black rectangles* and *short black rectangles*, respectively. Hydrophobic contacts are shown by *solid arrows*, hydrogen bonds and electrostatic contacts by *dashed arrows*, and contacts between protein backbone amides and DNA as *dotted arrows*. (For clarity, no distinction is made between hydrogen bonds and direct or indirect electrostatic interactions.) The identity of each nucleotide is shown in the center of its deoxyribose ring, and the protein residues are shown in italics. Intermolecular contacts are shown for (*a*) yeast TBP and the TATA element from the yeast CYC1(-52) promoter and *Arabidopsis thaliana* TBP and the TATA element of the adenovirus major late promoter; labels end with *y* or *a*, respectively (note that the bases of basepair 5 are represented as *hollow rectangles* because A or T is observed at this site); (*b*) the HMG box domain of SRY and the *MIS* gene consensus sequence; (*c*) the HMG box domain of LEF-1 and the TCRα gene consensus sequence; (*d*) IHF and the H′ site of phage 8 (the asterisk by T29-G30 indicates the site of the nick); (*e*) DBD2 of HMG I(Y) and a dodecamer comprising PRDII of the IFN β enhancer.

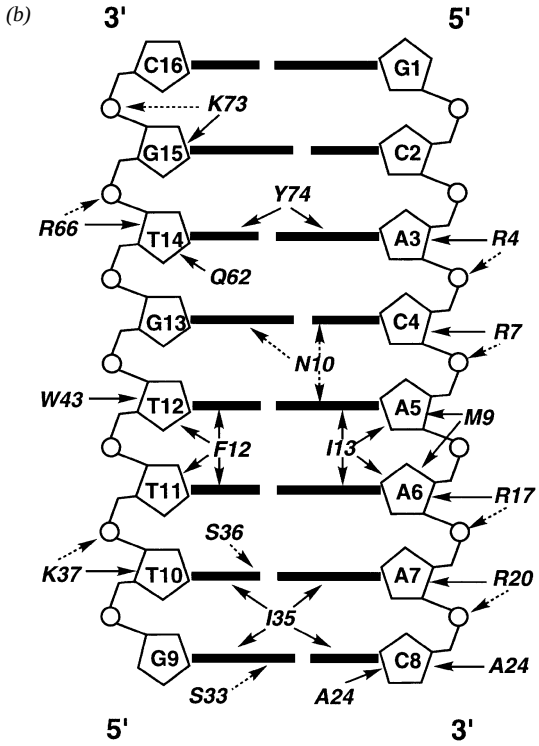

Figure 2 (*Continued*)

of the 16 hydrogen bond acceptors in the minor groove remain unsatisfied, with most of them being within van der Waals contact of nonpolar atoms. Upon binding, the TATA element is severely distorted, resulting in a widened and shallow minor groove. This distortion occurs largely by the insertion of a pair of phenylalanine residues located in the symmetrically displaced stirrups of TBP that insert into the first and last base steps within the 8 bp TATA box and kink the DNA (Figures 2*a*, 3*a*). In addition to the intercalating phenylalanines, the interface comprises eleven more hydrophobic residues that are within van der Waals contact of the minor groove bases (shown in the central portion of Figure 2*a*). With one exception, all of these interactions are symmetrical with respect to the pseudo-dyad of the complex and contact the central six base pairs of the TATA element; Pro191 contacts A1′ just outside of the upstream kink but the symmetrically displaced residue Ala100 makes no such contacts on the downstream side of the DNA.

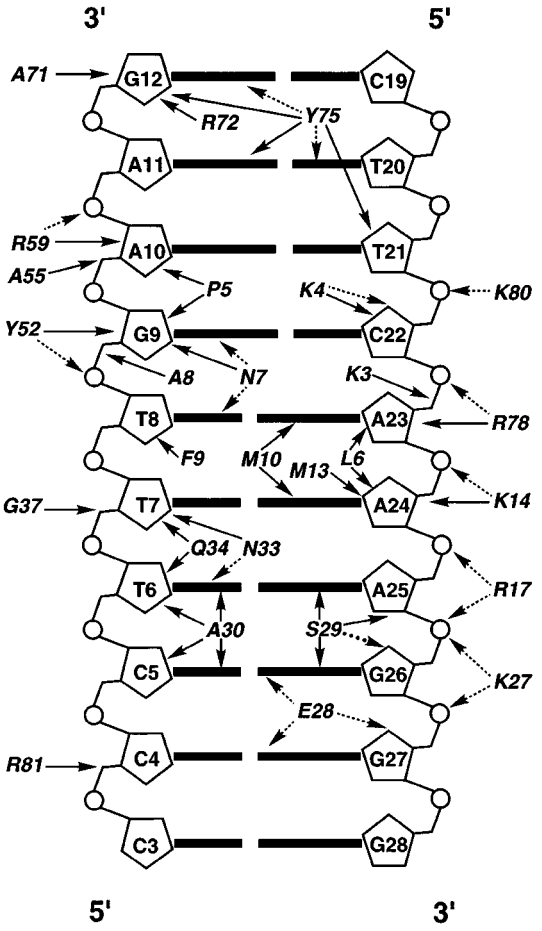

Figure 2 (*Continued*)

Only six hydrogen bonds are observed between the side chains of TBP and the base edges (Figure 2*a*). Asn159 and Asn69 are located in the center of the TATA box and contribute four hydrogen bonds to A4 and T5, and to T4′ and A5′, respectively. One base pair removed on either side, Thr215 and Thr124 hydrogen bond to N3 of A4 and A5′, respectively. Six direct contacts to the phosphate backbone from protein sidechains are seen in both crystal structures. They include contacts from the hydroxyl protons of Ser209 and Ser118 to the phosphate oxygens of A1′ and A8; and from the guanidinium groups of Arg98,

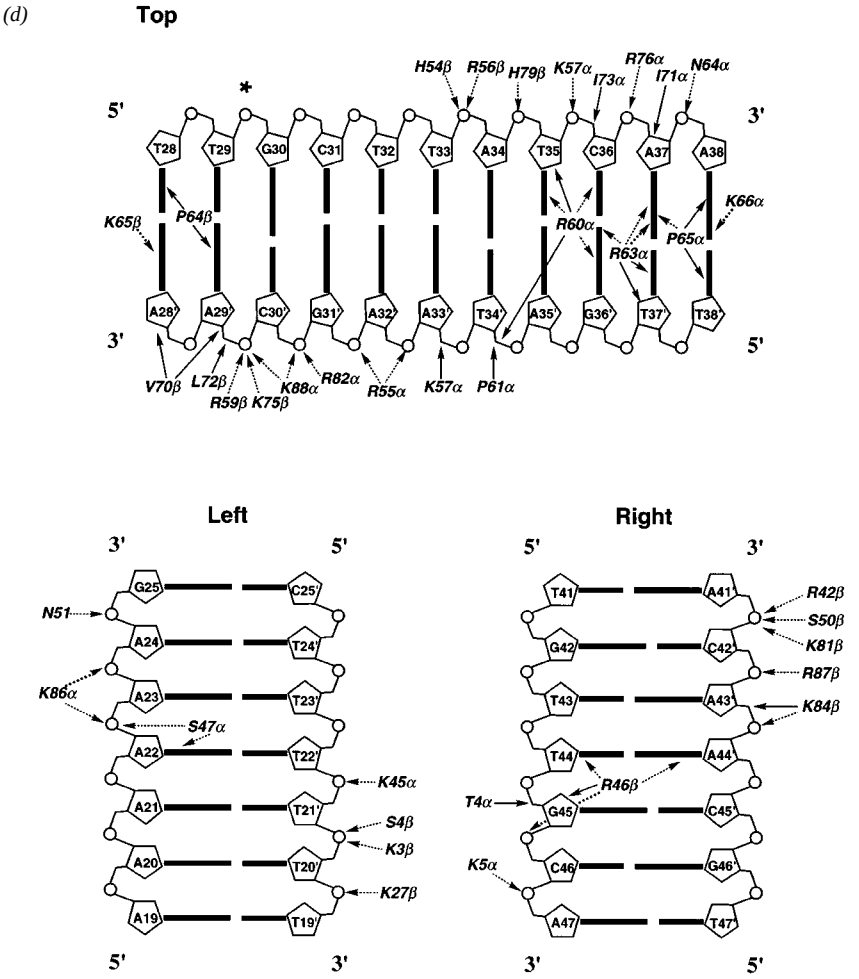

Figure 2 (*Continued*)

Arg105, and Arg196 to the phosphates near T8′, T6′, and T3, respectively. Located near to the phosphate backbone in both structures are several lysine residues; in the yeast TBP complex these lysine sidechains do not contact the DNA backbone either directly or indirectly, and are reported to be directed away from the backbone and stabilized by water-mediated intramolecular contacts (26). In contrast, in the structure of the *Arabidopsis* complex refined to 1.9 Å (24), lysines 85 and 176 contact the phosphate backbone directly, and lysines

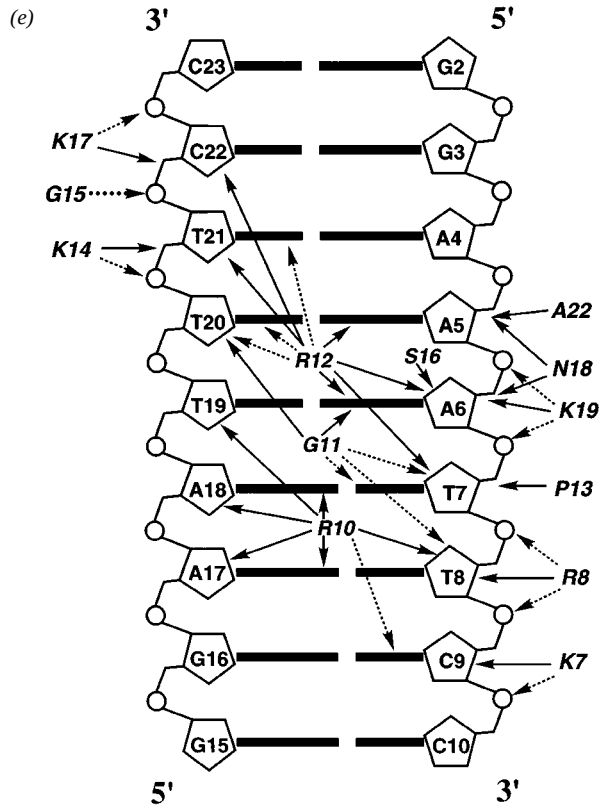

Figure 2 (*Continued*)

68, 78, 159, and 169 make water-mediated hydrogen bonds to the phosphate backbone. Despite the subtle differences in interactions seen in the two structures for the lysine sidechains, mutagenesis studies indicate that most of them are important for binding affinity, DNA recognition, or both (63).

When bound to TBP, the TATA element deviates significantly from canonical B-DNA and actually adopts several features characteristic of A-DNA, the most noteworthy being the extremely wide and shallow minor groove. Indeed, in the center of the TATA box the minor groove width exceeds 9 Å as compared to ∼4 Å for B-DNA (42). This prominent distortion results from an effective combination of interactions. The intercalating phenylalanine residues impart ∼45° kinks at the outer two basepair steps of the TATA box, and their nearly orthogonal orientations relative to one another cause buckling of the inner two basepairs

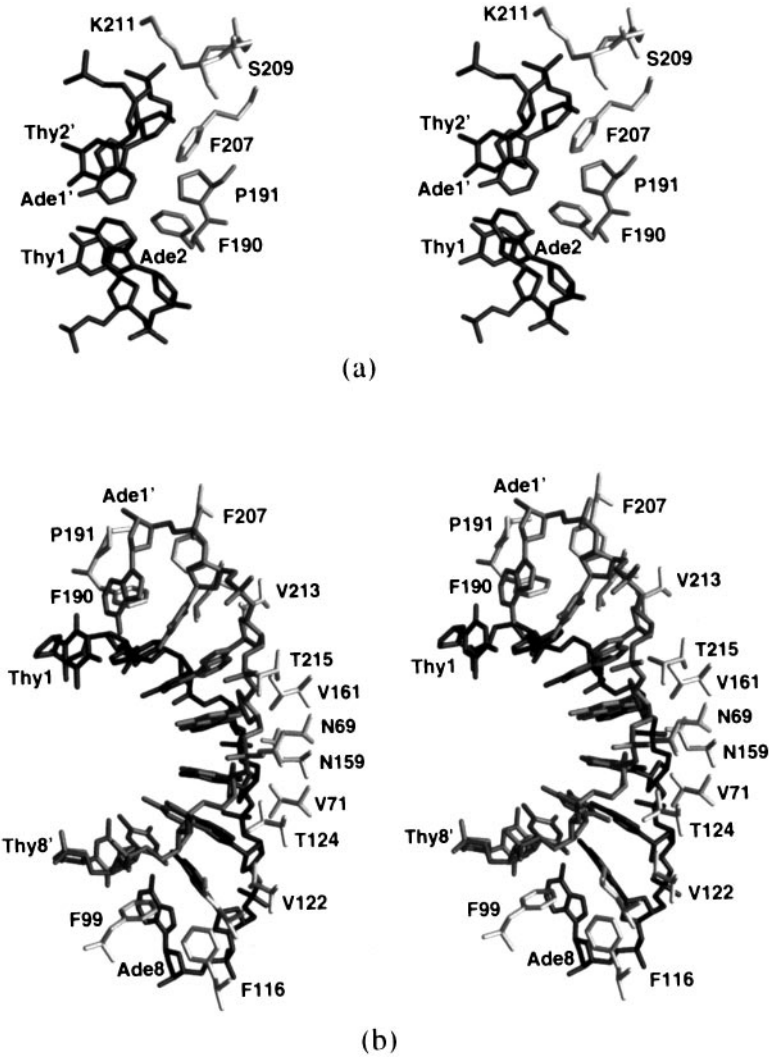

Figure 3 Stereoviews illustrating (*a*) hydrophobic residues in the stirrup of the C-terminal domain of yeast TBP inserted between the first basepair step T1-A1′/A2-T2′ of the TATA element; (*b*) the nearly symmetrical array of sidechains protruding from the concave underside of the saddle contacting the 8 bp TATA element.

which become unstacked (Figure 3*a*; rise = 5.4–5.9 Å). The hydrophobic interface coupled with the electrostatic interactions involving protein sidechains and the phosphate backbone induce a continuous and smooth bend of ~80° toward the major groove, as evidenced by the large average roll angle of +26° compared to ~0° for B-DNA. Although the DNA is underwound by 105° in only seven base steps (average twist angle = 21° for the TATA element as compared to 36° in B-DNA), this distortion is compensated for by one-third of a turn of positive supercoiling; thus, TBP exhibits no net effect on DNA supercoiling. Watson and Crick base-pairing is preserved throughout the DNA.

Architectural Factors That Contain HMG-Box Domains: The Male Sex Determining Factor SRY and Lymphoid Enhancer Binding Factor 1 (Lef-1)

Of the numerous architectural proteins that contain the ~80 residue DNA binding domains known as HMG boxes, the structures have been solved for only two in complex with DNA, namely SRY and LEF-1. The majority of HMG box-containing proteins bind nonspecifically to DNA, recognizing distinct structural features such as prebent or cruciform DNA rather than specific sequences (reviewed in 4, 18). SRY and LEF-1 are members of the high-mobility group proteins HMG-1/2 characterized by a single ~80 residue HMG-box domain (reviewed in 17). Although structurally distinct from TBP, SRY and LEF-1 also serve to bend DNA and do so in a manner analogous to TBP, namely by insertion of a hydrophobic amino acid between base steps from the minor groove side. Specific to preB- and T-cells, LEF-1 plays a requisite structural role in the formation of several multiprotein-DNA complexes, the best characterized of which is the T-cell receptor α (TCRα) gene enhancer (12, 13). Therein, LEF-1 occupies a centrally located site flanked by binding sites for at least three other transcription factors: downstream of LEF-1 the lymphoid specific transcription factors PEBP2 and ETS1 occupy adjacent and phased binding sites, and upstream of LEF-1 is an ATF/CREB binding site. ATF/CREB has been shown to stabilize the ternary nucleoprotein complex containing PEBP2 and ETS1 and to interact with PEBP2 in vitro. Moreover, mutations in the TCRα enhancer that alter the relative spatial relationship of any of these proteins impair enhancer function. These findings are consistent with the formation of a multiprotein-DNA complex mediated by DNA bending by LEF-1. The male sex-determining factor SRY likely plays an analogous role in transcriptional activation of the genes important in sexual development such as the Müllerian inhibiting substance (*MIS*) gene, whose product accounts for the regression of the female Müllerian ducts in male embryos (14, 60). Mutations in SRY are responsible for 15% of male to female sex reversal (46XY females).

SRY The structure of a complex of the 78 amino acid HMG box domain of SRY (referred to as SRY below) with a DNA octamer comprising its target site in the *MIS* promoter (Table 1) has been solved by NMR (59). SRY has a twisted L shape that comprises irregular N- and C-terminal strands that lie adjacent to one another and three α helices (Figure 1). The long arm of the L is formed by helix 3 and the N- and C-terminal strands, and the short arm of the L is formed by helices 1 and 2. Orientation of the helices is maintained by three concentrated regions of hydrophobic packing located at the center of the long arm, the tip of the short arm, and the corner of the L. The minor groove binding surface consists of helices 1 and 3, and is bordered at the bottom by a ridge comprising helix 2 and at the top by a ridge comprising the opposing terminal strands. Providing another example of induced fit, the *MIS* sequence is significantly distorted upon SRY binding [NMR studies show the free *MIS* DNA to be essentially B-form (57)] to yield a widened and shallow minor groove whose convex surface is molded perfectly to the concave binding surface of SRY.

Numerous hydrophobic and electrostatic interactions contribute to the severe DNA distortion in the SRY complex, with the most apparent deformations located near the center of the octamer at basepairs 5 and 6, and at the bottom of the octamer at base steps 2 and 3 (Figure 4). At the center, Met9, Phe12, Ile13 and Trp43 form a hydrophobic wedge that is anchored to basepairs 4 and 5 by electrostatic interactions involving Asn10 (Figure 4*b*). At the center of the wedge Phe12, which is positioned orthogonally to the bases of the minor groove, and the intercalating Ile13 both interact with the bases to induce a large bend toward the major groove, while Met9 and Trp 43 on the outer sides of the wedge contact the ribose rings to pry open the minor groove. The bend at base steps 2 and 3 occurs as Tyr74 packs against A3 and T14, anchoring SRY to the top of the DNA (Figure 4*a*); and Ile35 contacts base pairs 7 and 8 to anchor SRY to the bottom of the DNA (Figure 4*c*). The sidechains of an additional seven basic residues make electrostatic contacts with the phosphodeoxyribose backbone to provide an extensive scaffold for DNA recognition (Figure 2*b*).

Upon binding to SRY, profound structural changes are induced in the *MIS* DNA as it leaves the essentially B-form conformation to take on structural features comparable to those of A-DNA; indeed, the rms difference between the SRY-bound DNA and classical A-DNA is 2.4 Å as compared to an rms difference of 4.2 Å to classical B-DNA. The minor groove is significantly expanded in the complex as revealed by the average width of 9.4 Å; and the DNA is underwound exhibiting an average interbase helical twist of $26^\circ \pm 6^\circ$, and an average interbase pair rise of 4.1 ± 0.3 Å, which exceeds that seen in both A-DNA (3.4 Å) and B-DNA (3.6 Å). Six of the seven base steps display positive roll angles, ranging from 10.5° between base pairs 7 and 8 to a maximum of 35° between base pairs 2 and 3, to produce an overall bend angle of $\sim 70^\circ$–80°,

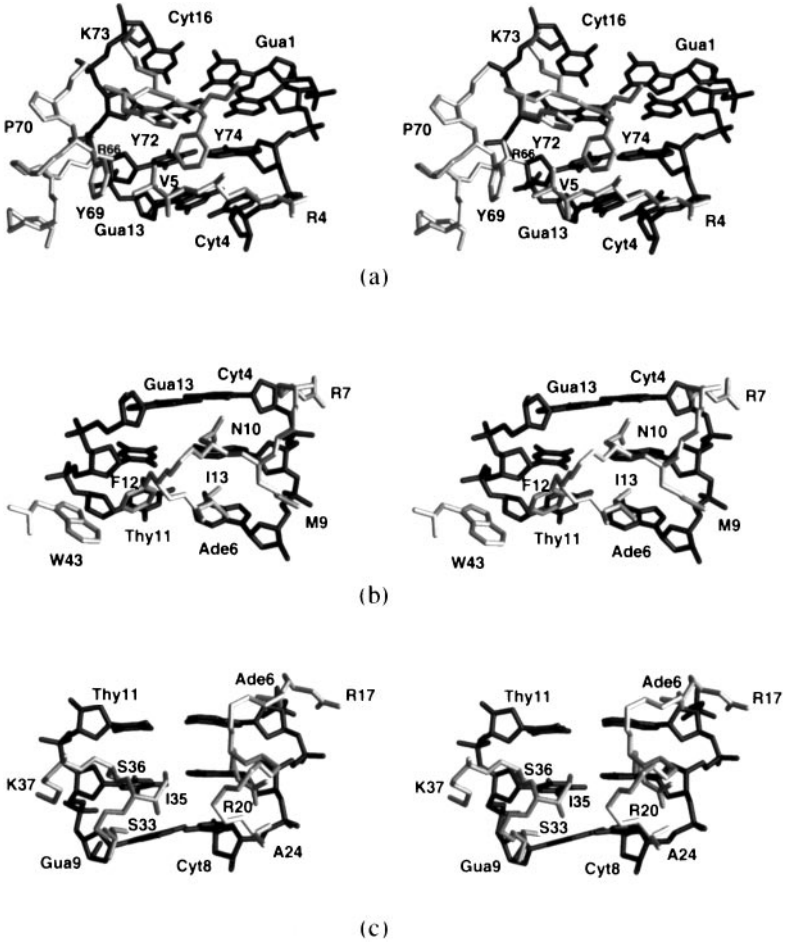

Figure 4 Stereoviews illustrating interactions between hSRY-HMG and DNA. These include (*a*) sidechains at the C-terminal end of helix 3 and the C-terminal and N-terminal strands; (*b*) the clustering of hydrophobic residues at the center of the octamer with partial intercalation of Ile13; (*c*) sidechains from helices 1 and 2 near the 3′ end of the coding strand of DNA.

and all but two of the ribose rings exhibit sugar puckers indicative of A-DNA (O1′-endo to C3′-endo).

LEF-1 The NMR structure of LEF-1 complexed to a 15 bp oligonucleotide from the TCRα enhancer (Table 1) is very similar to that of the SRY complex (30). LEF-1 exhibits the twisted L shape of the minor-groove binding HMG domains, and the wide and shallow DNA is bent by $\sim 117^{\circ}$ toward the major groove. LEF-1 uses a methionine (Met10) residue rather than an isoleucine for

intercalation, which is stabilized by hydrophobic contacts from Phe9 and Met12 to nearby bases. An additional difference between the two structures is seen in the length and interactions of the C-terminal strands where the construct used in the LEF-1 structure contains five additional amino acids, four of which are lysines. Thus, the C-terminal tail extends across the compressed major groove to make electrostatic contacts to the phosphate backbone of the DNA from the major groove side. These additional contacts may account for the larger bend angle seen in the LEF-1 complex.

Integration Host Factor

Integration host factor (IHF) is a prokaryotic architectural protein that takes DNA bending to a new extreme: IHF causes the DNA to adopt a U-turn (39). The biological roles in which IHF has been implicated are numerous and varied, but the best characterized is its role in site-specific recombination in bacteriophage λ (16, 31). In this system, several specific IHF binding sites occur near the phage attachment (*att*) sites, which participate either in integrative recombination or excisive recombination. IHF comes into play because its main function is to bend the DNA so as to bring into proximity the distant promoter binding sites of the heterobivalent λ integrase. Indicative of IHF's role as a classical architectural protein, λ recombination can occur when IHF's binding sites are replaced by intrinsically curved DNA or by binding sites of other architectural proteins that also serve to bend DNA (48). Specifically, recombination has been demonstrated where IHF was replaced with the closely related nonspecific DNA-binding protein HU, the otherwise unrelated architectural proteins HMG-1 and HMG-2, and the eukaryotic histone dimer H2A-H2B.

Interestingly, IHF appears to play a dual role in the initiation of replication: Besides its defined structural role, IHF acts as a regulatory protein in concert with FIS [factor for stimulation of inversion (5)], another site-specific DNA-bending protein. In a more traditional sense, IHF can function as a transcriptional activator or a repressor, and repression can occur via two different mechanisms: Namely, IHF can stabilize binding of a given repressor (1) or it can occlude binding of an activator by directly occupying its specific binding site. Thus, IHF's multifaceted role arises from its ability to bind sequences specifically as well as to bend DNA.

IHF is a heterodimeric protein composed of two ∼10 kDa homologous subunits referred to as α and β. The two subunits intertwine to produce a unique protein fold that resembles a body with two long arms reaching out to wrap around the DNA (Figure 1). The bottom of the body is made up of two helices per subunit that are packed orthogonally against their equivalent in the second subunit, and the top of the body is formed by a third short helix and two antiparallel β-sheets from each subunit. Extending away from the top of the body

are the two arms, each comprised of two antiparallel β-sheets. Together, the β-sheets of the top of the body and the arms reach out to "embrace" the DNA, continuously following the minor groove so that binding occurs concomitantly on opposite faces of the DNA.

The crystal structure of IHF complexed to a 35 bp oligonucleotide from the H′ site of phage λ (Table 1) shows that the DNA is bent by $\sim 160^{\circ}$ (Figure 1). Although the overall structure of the complex appears to be symmetrical, the opposing DNA and protein sequences and their detailed interactions are asymmetrical. Like other IHF consensus binding sites, the H′ site contains three characteristic sequences, namely two conserved elements, 5′-TATCAA-3′ and 5′-TTG-3′, and a 6-bp A-tract. The center of the U-turn is positioned at the 5′ end of the TATCAA consensus site, and the dA/dT-rich sequence and the TTG consensus are located one helical turn 5′ and 3′ to the start of the TATCAA element, respectively. Together, the contacts between the DNA and the arms and body of the protein bury in excess of 4500 Å^2 of surface area accessible to solvent in the uncomplexed molecules.

Analogous to TBP, SRY, and LEF-1, IHF employs two hydrophobic residues, in this case Pro65α and Pro64β, located at the tips of opposing arms, to introduce substantial bends in the DNA. These proline residues, which are absolutely conserved among all members of the IHF and HU families, intercalate between base steps separated by nine basepairs making extensive hydrophobic contact with the DNA bases (Figures 2*d*, 5*a*).

In addition, hydrogen bonds between the prolines and N3 of both adenines 5′ to the kinks are observed. The only other hydrophobic interactions occur between hydrophobic sidechains and deoxyribose rings, as opposed to minor groove bases. These involve contacts between the pseudosymmetrically displaced Ile71α/Val70β and Ile73α/Leu72β, and deoxyribose rings nearby the kinks, and contacts from Pro61α to the sugar of T-34. In the crystal structure, the α arm of IHF (Figure 5*a*) reaches deep into the minor groove of the consensus sequence 5′-TATCAA-3′ as the aliphatic protons of Arg63α and Arg60α pack against basepairs 34–37 and the backbone and sidechains make hydrogen bonds to the same basepairs. On each side of the complex, the N-terminus of helix 1 and the $\beta 1\beta 2$ loop of one subunit join with the N-terminus of helix 3 of the second subunit to form a "clamp" rich with electrostatic contacts to the DNA backbone. At the tip of the $\beta 1\beta 2$ loops, Ser47α and Arg46β insert into the minor groove on the left (Figure 5*b*) and right (Figure 5*c*) sides of the complex, respectively. The backbone of helices 1 and 3, located below and above the $\beta 1\beta 2$ loop, contact phosphates located on opposite sides of the minor groove. In addition to the intermolecular interactions centered around the tips of the arms and the sides of the body of the protein, numerous positively charged sidechains contact the phosphate backbone of the DNA, contributing to the overall bend.

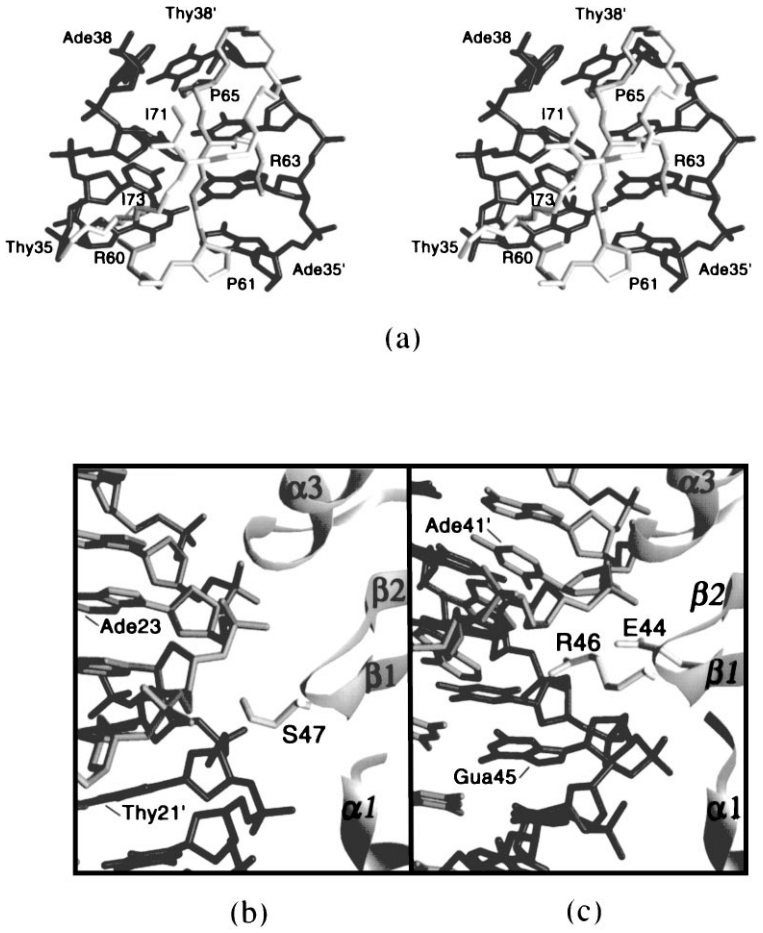

Figure 5 (*a*) Stereoview illustrating contacts between the α arm of IHF and the first DNA consensus sequence 5′-TATCAA-3′ (T33-A38); (*b*) closeup of the left side of the protein body contacting the A-tract (A19-A24); (*c*) closeup of the right side of the protein body contacting the second DNA consensus sequence T43-T44-G45 [adapted from (39)]. Labels for the α-helices and β-strands of IHFβ are italicized.

The arrangement of protein-DNA interactions seen in the IHF complex, in combination with DNA sequences that are susceptible to deformation, introduce an alternating and phased pattern in the DNA parameters, most notably in the minor groove width (39) (Figure 1). The intercalating prolines introduce short stretches of very wide minor grooves (>10 Å) located approximately one helical turn apart at the top corners of the bent DNA. These are in turn evenly interspersed with three distinct regions of narrowed minor groove, each of which

is separated by one helical turn (∼10 basepairs), and thus located at the center and sides of the inside of the DNA bend. Thus, the minor groove incorporating the three consensus elements alternates between narrow and wide every half helical turn, and the three sites possessing the narrowest minor grooves are positioned to contact the top and sides of the compact body of the protein. This arrangement accommodates the abundance of direct and water-mediated electrostatic interactions between the body of the protein and the narrow minor grooves. The DNA conformation returns to B-form just outside of the first and last consensus elements. The kink introduced by the prolines is characterized by an unprecedented 57° roll angle, and the widened minor groove surrounding the kinks continues 2 to 3 base steps in either direction. The regions of narrowed minor grooves range from ∼2.5–3.5 Å in width.

High Mobility Group Protein I(Y)

The high mobility group I (HMG I) family of proteins, which includes HMG I, its isoform HMG Y, and HMG I-C, are non-histone chromosomal proteins that are entirely distinct from the other HMG box HMG 1/2 family and the HMG 14/17 family (reviewed in 4). Members of the HMG I family are basic proteins ∼10 kDa in size that consist of a variable N-terminus, an acidic C-terminus, and three short DNA binding domains (DBDs) that are separated by linkers of 11 to 23 amino acids. HMG I(Y) binds to the minor groove of AT-tracts with nanomolar affinity and appears to use two DBDs for binding, which explains the recurrence of two helically phased AT-tracts in naturally occurring enhancers. Analogous to the architectural proteins described above, HMG I(Y) alone does not activate transcription but plays a critical architectural role in the assembly of enhanceosomes on what appears to be a growing number of cytokine and viral genes (10, 22, 55). Further, HMG I(Y) has been shown to interact directly with transcriptional activators of the Rel, bZIP, ETS, and POU families, stabilizing their binding to DNA. In contrast to the previously described architectural proteins, HMG I does not induce bends in DNA, but instead acts to preserve the B-form character of DNA, and can actually reverse moderate bends in DNA (9). The role of HMG I in facilitating the assembly of a large nucleoprotein structure has been best characterized for the complex, yet compact (∼70 bps, all of which are essential for viral induction) interferon β gene enhancer in which HMG I helps to recruit the transcriptional activators NFkB, ATF2/cJun and IRF (55).

The NMR structure of a truncated form of HMG I (residues 50–91) that contains DBDs 2 and 3 complexed to two equivalents of a dodecamer comprising the PRD2 site of the IFN β gene enhancer revealed two distinct DNA binding motifs (21). The high affinity DBD2 (Figure 1) consists of three modular components comprising the extended Arg-Gly-Arg core sequence that inserts deep into the

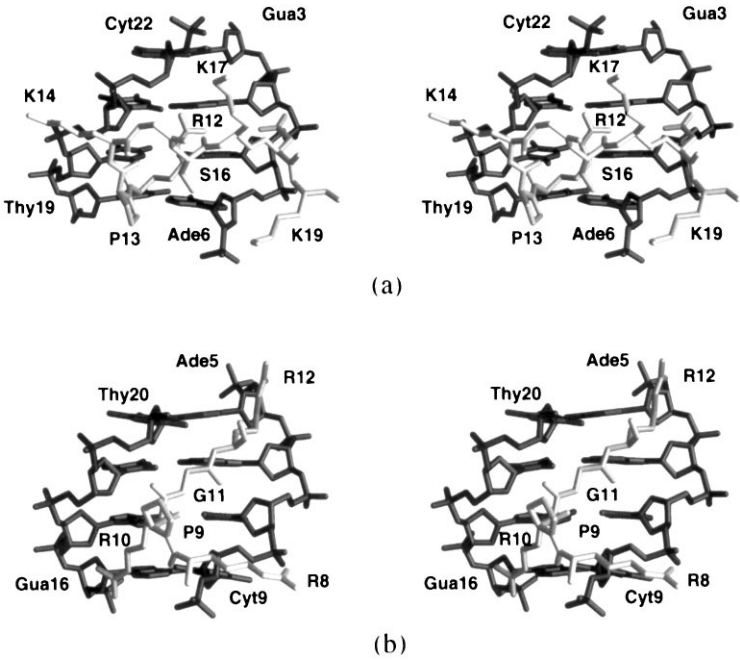

Figure 6 Stereoview illustrating contacts between the high-affinity binding domain DBD2 of HMG I(Y) and (*a*) the 5′ end of the AT-tract, and (*b*) the 3′ end of the AT-tract of PRDII in the IFN β enhancer.

minor groove, a pair of lysine and arginine residues that flank the core and make hydrophobic and electrostatic interactions to the sugar-phosphate backbone, and a polar network of six amino acids C-terminal to the basic core. The four residues directly following the Arg-Gly-Arg core (Pro13, Lys14, Gly15, and Ser16) form a type II β-turn that effectively bridges the minor groove, and the remaining three residues clamp around the lower strand of the DNA using polar and nonpolar contacts (Figure 6*b*). The lower affinity DBDs 1 and 3 (not shown) lack the additional six amino acids, which accounts for their ~100-fold decrease in DNA binding affinity.

The high affinity DNA-binding motif (DBD2) of HMG I is remarkable in that it contains only 13 amino acids that upon DNA binding undergo a conformation change from a random coil (as shown by NMR studies of free HMG I) to a fully structured motif in which every amino acid, with the exception of one proline, contacts the DNA (Figure 2*e*). The Arg-Gly-Arg core lies deep in the minor groove, with the glycine centered between base pairs 6 and 7 and the arginine

sidechains extending in either direction along the floor of the minor groove such that every H2 atom of the adenines is encountered by either Arg10 or Arg12 (Figures 2*e*, 6*a*, 6*b*). The guanidinium groups of Arg10 and Arg12 are hydrogen-bonded to the O_2 atoms of C9 and T21, respectively. At the top of the β-turn, Lys14 and Gly15 contact the phosphoribose backbone of the top DNA strand, and Pro13 and Ser16 of the sides of the turn contact the ribose rings of the bottom DNA strand at T7 and A6, respectively (Figure 6*a*). The backbone of the final three amino acids is oriented orthogonal to the DNA axis and the sidechains extend across the groove again to contact the sugar-phosphate backbone on both sides of the minor groove (Figures 2*e*, 6*a*).

The conformation of the PRD2 dodecamer complexed to HMG I is essentially B-form. The average values of 3.6 Å for the local helical rise, 35.6° for the local helical twist angle, and ∼1° for the inter-base pair roll angles are all indicative of classical B-form DNA. In addition, the average minor groove width is only ∼1–1.5 Å larger than that seen in B-DNA.

FACTORS CONTRIBUTING TO SPECIFIC BINDING AND CONFORMATIONAL CHANGES

Recognition

In contrast to transcriptional activating factors that tend to recognize very specific DNA sequences and features of the major groove, the minor groove architectural proteins are faced with the problem of specific recognition in the barren minor groove (47). Following the determination of the structures of five architectural proteins bound to their DNA targets, it is becoming clear that several strategies are used for recognition and alteration of DNA conformation. These include exploitation of sequence-dependent DNA deformability, recognition of narrower than usual minor grooves, and asymmetric charge neutralization along the sugar-phosphate backbone (50). Of those proteins described above that induce extreme bends, partial insertion or intercalation of hydrophobic residues imparts large positive roll angles (i.e. toward the major groove) that contribute greatly to the overall bending of the DNA (58). Thus far, insertion of these residues occurs at YR or RR/YY steps (where "R" stands for purine and "Y" for pyrimidine) which have been shown in naked duplex DNA to be either inherently bent or subject to deformations (15, 51). In the case of TBP, the phenylalanines of the C-terminal domain insert between the poorly stacked and easily distortable T1-A2 step of the TATA box and impart a greater positive roll angle than is seen at the Y7-Y8 step at the opposite end of the 8 bp TATA element. In SRY, Ile13 is inserted at the RR step A5-A6. Although the RR/YY steps do not appear to be as readily distortable as the YR steps, they still display

large deviations in parameters basic to DNA bending, namely roll angle and degree of twist (53). The most obvious differences between YR and RR/YY steps is the degree of base stacking, and for AA/TT sequences the relative positions of the thymine methyl groups. Thus, it follows that in cases where positive roll angles at AA/TT steps are large enough to disrupt base stacking and impose steric clash among the thymine methyls, these energetically costly transformations must be compensated for by the more complete packing or pi-pi interactions that can occur among the sidechains inserted between sequential and similar bases than among alternating (YR or RY) base steps.

Both HMG I(Y) and parts of IHF bind to B-form AT-rich DNA. In the case of HMG I(Y), the crescent formed by the extended Arg-Gly-Arg sequence inserts deeply into, and runs along the floor of, the minor groove, taking advantage of the lack of irregular YR steps. In the case of IHF, the top and both sides of the body of the protein interact with the three regions of unusually narrow minor grooves. Clearly, in two of these interactions, namely at the A-tract and the center of the DNA, the narrow minor grooves are important for recognition considering that binding occurs through water-mediated contacts with the minor groove spine of hydration that persists upon binding. Because IHF induces a nearly 180° turn in its bound DNA, it is only appropriate that this protein should use all three strategies for deformation: The body of IHF is positively charged and lies on the inside of the bent DNA to neutralize the phosphate backbone and contribute to the overall bend.

Specific and Directional Binding

With each new structure of a protein-DNA complex comes the question of how a particular protein recognizes a specific DNA sequence. In the case of TBP, IHF, and HMG I, yet another complexity arises resulting from the fact that each of these proteins uses pseudo-symmetric motifs to interact with DNA. Moreover, they are capable of recognizing more than one sequence while still conserving directional binding. Several thoughtful analyses of TBP-TATA box recognition have appeared recently and provide insight into the role that protein and DNA structure and sequence-specific deformability play in influencing directional binding. In a revealing study by Suzuki et al (52), the coordinates of a cocrystal structure were duplicated and one of the identical structures was rotated 180° about the pseudo-dyad. The resulting oppositely oriented molecules were best fit to β-strands 1 and α-helices 2 located at the center of the dimer. [This exercise was performed for both the Sigler (26) and Burley (25) structures.] These models clearly show that in both structures the β-sheet of the C-terminal domains, which bind to the 5′-TATA-containing side of the DNA, display greater curvature and thus approach more closely to the DNA than the β-sheet of the N-terminal domains which bind to the 3′-A–rich half of the DNA. In turn, the

5′-TATA ends of the DNA display greater curvature and compression of the major groove than the 3′-A–rich sequence. A residue-by-residue comparison of the two protein domains reveals that for the β-sheet residues directed away from the DNA, those in the C-terminal domain are larger than those in the N-terminal domain. [This finding holds true for the uncomplexed TBP structures as well, indicating that the greater curvature is inherent in TBP itself. (6, 33)] Increased curvature of the protein backbone away from the protein interior on the C-terminal side alleviates the greater steric crowding and coincidentally increases curvature toward the DNA. Conveniently, the 5′-TATA side of the DNA can accommodate the increased curvature of the C-terminal domain better than the 3′-A tract because it contains alternating thymines and adenines that do not have to contend with steric clash between the methyls of the thymines in the major groove. Indeed, measurement of the distances between methyls of the A-rich side show that they are at the limit of major groove compression. Also contributing to the greater curvature of the C-terminal side of TBP is the strictly conserved and asymmetrically-positioned proline toward the end of helix two, which causes a slight helical bend directed toward the underlying saddle. In addition to the asymmetric and complimentary curvatures of the protein and DNA, Dickerson and coworkers (23) have noted that a single proline (191y/149a) that packs against the 3′ side of A1′ may be largely responsible for the preference for T over A as the first site, considering that substitution to AT would introduce unfavorable interactions between atoms of the proline ring and the C2 oxygen of thymine.

It had been proposed that asymmetrical charge distribution might play a role in orienting TBP on the DNA (26). However, Kim and Burley (24) noticed that any considerable asymmetry in charge distribution that would affect DNA binding is limited to the stirrups where Arg55y and Arg56y in the C-terminal domain are replaced by Glu144y and Glu146y in the N-terminal domain. With the exception of Arg56, the sidechains of these residues are directed away from the DNA and probably do not appreciably affect binding polarity. This view was substantiated by the ternary crystal structure of TBP/TFIIB/TATA element that shows Glu144 and Glu146 to be hydrogen bonded to TFIIB residues rather than interacting with the DNA (35). [More recently the ternary crystal structure of a complex comprising an archaeal protein homologous to TBP that recognizes so-called box A motifs, a TFIIB homolog, and a box A DNA target has been solved (27); in this case, charge distributions distinct from those seen in eukaryotic TBPs play a strong role in orientation (8, 27).]

HMG I functions differently from the pseudo-symmetrical DNA-bending proteins TBP and IHF because it stabilizes B-form DNA and can actually reverse moderate bends (9). The core DNA-binding motif of HMG I comprises the palindromic sequence Pro-Arg-Gly-Arg-Pro, which binds to AT-rich sequences,

preferably of 5 or 6 base pairs in length. The solution structures of the high- and low-affinity domains of HMG I bound to the sequence 5′-GGGAAATTCCTC-3′ showed that both domains are oriented in a single direction (21). Arg10 is in close contact with the H2 atoms of A17 and A18 on the left strand of the DNA, while Arg12 contacts the H2 atoms of A5 and A6 on the opposite strand (Figure 6*a*). Thus, the extended concave surface of Arg-Gly-Arg inserts deeply into the minor groove and runs parallel with the right handedness of it. If the orientation were reversed, there would be steric clash with the O_2 atoms of the complementary thymine. In regard to specificity, the sequence 5′-AAATT-3′ of the PRD II element lacks any disruptive YR steps, and the structure of HMG I complexed to DNA shows that the narrower and more regular the minor groove is, the more extensive are the van der Waals contacts. The contacts will be further augmented by the sequences AAa/tT or AAa/ta/tT, which place the H2 atoms on opposing strands as far apart as possible; that is, they run with the floor of the minor groove. If the DNA sequence were reversed, for example to TTTAA, the H_2 atoms would be positioned more closely atop one another, that is, running more perpendicular to the floor of the groove, which would preclude the extensive van der Waals contacts seen in the structure by virtue of a decreased distance and steric hindrance of the O_2 atoms of the thymines. Further, the sequence would also include an RY (in this case TA) step, which would likely disrupt the straight and narrow groove of the DNA.

HMG I has been reported to be a non-specific binder of AT-rich sequences. It remains possible that it can bind to the less favorable TTAA-containing sequences, given the flexibility of the Arg-Gly-Arg motif and the availability of two more binding motifs that can be used to enhance the overall DNA-binding affinity. We may also find that the orientation is affected in the presence of interactions with other transcriptional activators.

As with TBP and HMG I, the pseudosymmetrical IHF binds in a directional manner. However, compared to TBP and HMG I, IHF is considerably less symmetrical in both protein and DNA sequence at the sites of intermolecular contacts. There are four main areas of protein-DNA interaction in the cocrystal structure of IHF: These include the two sides of the protein body interacting with the narrowed minor grooves on the left and right side of the DNA and the two arms reaching into the upper left and right curves of the U-turn (Figures 5*a–c*). Further, the top of the protein body and the center of the DNA are connected by a matrix of ordered water molecules that originates as the spine of hydration in the narrowed minor groove of this region. The two conserved consensus sequences are conveniently located on the 3′ or right half of the DNA such that together they likely direct the orientation of IHF on the DNA. Given the degree of bending and the extent of complementary surfaces, direct contacts between IHF and the edges of the DNA bases are scant, but a closer look at the interactions about the

consensus sequences offers some explanation for specificity and recognition. On the right side of the complex at the second DNA consensus sequence, Arg46β is rigidly positioned, through the support of salt bridges with proximal sidechains, in the center of the groove between basepair steps T-A 44 and G-C 45 (Figure 5*c*). Arg46β thus contacts the deoxyribose rings on the top and bottom strands of the DNA. Rice et al (38, 39) note that at this site the two sides differ. On the consensus side, helix 1 of the α subunit is recessed toward the body of the protein, and this gap conveniently accommodates the protruding phosphate backbone that results from the over-twisted and slightly kinked YR step T44-G45. On the left side, the A-tract produces a uniform and narrow groove that maintains a spine of hydration through which the protein contacts the DNA indirectly (Figure 5*b*). This feature could account for the relaxed specificity observed for this site.

The interactions between the arms of IHF and DNA differ in that the α arm, which interacts with the first DNA consensus site, sits more deeply in the minor groove than the β arm. The most obvious differences in the two sides are that Arg60α and Arg63α make direct contacts to conserved bases, while their equivalents in the β arm, Arg59β and Arg62β, do not directly contact the DNA (Figure 2*e*). This discrepancy may be real or it may be an artifact of the complex as the DNA is nicked between T29 and G30, and non-Watson-Crick base pairing results as the surrounding bases are involved in crystal packing.

CONCLUSIONS

The structures of the five architectural proteins reviewed above have given us an integral view of the variety of ways proteins recognize and remodel DNA conformation purely through intermolecular interactions occurring in the minor groove. All of these proteins play crucial roles in the assembly of very large protein-DNA complexes, such as the intasome (40) or enhanceosome (55), that are necessary for carrying out myriad cellular processes. While the vast majority of "non-architectural" proteins bind to the major groove of DNA, most of the architectural proteins interact with the minor groove, and their precisely spaced binding sites coupled with the degree of bending (or unbending) allows for the controlled assembly of these macromolecular complexes (56).

Just as the initial structures of TBP in complex with TATA elements led the way for larger ternary complexes (11, 35, 54), we can expect the same progression for the other architectural factors as we await ternary (or larger) structures involving IHF, LEF-1, SRY, and HMG I(Y). On the other hand, we still look forward to seeing high-resolution structures of protein-DNA complexes for those architectural proteins that recognize specific DNA structures, such as members of the HMG 1/2 family that bind to prebent or cruciform DNA, and proteins

involved in chromatin structure. With the continuing advances in stably modifying DNA structure, whether it be through covalent modifications to force specific degrees of bends (61, 62), or through methods to stabilize particular conformation such as the cruciforms basic to recombination (29, 46), we can anticipate structures for these nonspecific binders as well.

ACKNOWLEDGMENTS

GMC and AMG acknowledge the support of the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health for experimental work in the authors' laboratory. CAB acknowledges a Cancer Research Institute Postdoctoral Fellowship.

Visit the *Annual Reviews home page* at http://www.AnnualReviews.org.

Literature Cited

1. Betermier M, Rousseau P, Alazard R, Chandler M. 1995. Mutual stabilization of bacteriophage Mu repressor and histone-like proteins in a nucleoprotein structure. *J. Mol. Biol.* 249:332–41
2. Buratowski S, Hahn S, Guarente L, Sharp PA. 1989. Isolation of the gene encoding the yeast TATA binding protein TFIID— a gene identical to the spt15 suppressor of ty element insertions. *Proc. Natl. Acad. Sci. USA* 86:5718–22
3. Burley SK, Roeder RG. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65:769–99
4. Bustin M, Reeves R. 1996. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* 54:35–100
5. Cassler MR, Grimwade JE, Leonard AC. 1995. Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J.* 14:5833–41
6. Chasman DI, Flaherty KM, Sharp PA, Kornberg RD. 1993. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl. Acad. Sci. USA* 90:8174–78
7. Cormack BP, Struhl K. 1992. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* 69:685–96
8. DeDecker BS, O'Brien R, Fleming PJ, Geiger JH, Jackson SP, et al. 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* 264:1072–84
9. Falvo JV, Thanos D, Maniatis T. 1995. Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* 83:1101–11
10. Farnet CM, Bushman FD. 1997. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* 88:483–92
11. Geiger JH, Hahn S, Lee S, Sigler PB. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* 272: 830–36
12. Giese K, Kingsley C, Kirshner JR, Grosschedl R. 1995. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* 9:995–1008
13. Giese K, Pagel J, Grosschedl R. 1994. Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. *Proc. Natl. Acad. Sci. USA* 91:3368–72
14. Goodfellow PN, Lovell-Badge R. 1993. SRY and sex determination in mammals. *Annu. Rev. Genet.* 27:71–92
15. Goodsell DS, Kaczor GM, Dickerson RE. 1994. The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J. Mol. Biol.* 239:79–96
16. Goosen N, van de Putte P. 1996. The regulation of transcription initiation by integra-

tion host factor. *Mol. Microbiol.* 6:2557–63
17. Grosschedl R. 1995. Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell Biol.* 7:362–70
18. Grosschedl R, Giese K, Pagel J. 1994. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet.* 10:94–100
19. Hernandez N. 1993. TBP, a universal transcription factor? *Genes Dev.* 7:1291–1308
20. Hoffmann A, Sinn E, Yamamoto T, Wang J. 1990. Highly conserved core domain and unique N-terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature* 346:387–90
21. Huth JR, Bewley CA, Nissen MS, Evans JN, Reeves R, et al. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* 4:957–65
22. John S, Reeves RB, Lin JX, Child R, Leiden JM, et al. 1995. Regulation of cell-type-specific interleukin-2 receptor alpha-chain gene expression: potential role of physical interactions between Elf-1, HMG-I(Y), and NF-kappa B family proteins. *Mol. Cell Biol.* 15:1786–96
23. Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, et al. 1996. How proteins recognize the TATA box. *J. Mol. Biol.* 261:239–54
24. Kim JL, Burley SK. 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* 1:638–53
25. Kim JL, Nikolov DB, Burley SK. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520–27
26. Kim Y, Geiger JH, Hahn S, Sigler PB. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512–20
27. Kosa PF, Ghosh G, DeDecker BS, Sigler PB. 1997. The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. USA* 94:6042–47
28. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, et al. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–54
29. Lilley DMJ, Clegg RM. 1993. The structure of branched DNA species. *Q. Rev. Biophys.* 26:131–75
30. Love JJ, Li X, Case DA, Giese K, Grosschedl R, et al. 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 376:791–95
31. Nash H. 1996. The HU and IHF proteins: accessory factors for complex protein-DNA assemblies. In *Regulation of Gene Expression in* Escherichia coli, ed. ECC Lin, AS Lynch, pp. 149–79. Austin, TX: RG Landes
32. Nicholls AJ, Sharp K, Honig B. 1991. Protein folding and associations: insight from interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Func. Genet.* 11:281–96
33. Nikolov DB, Burley SK. 1994. 2.1-Å resolution refined structure of a TATA box-binding protein (TBP). *Nat. Struct. Biol.* 1:621–37
34. Nikolov DB, Chen H, Halay ED, Hoffman A, Roeder RG, et al. 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA* 93:4862–67
35. Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, et al. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377:119–28
36. Pabo CO, Sauer RT. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053–95
37. Peterson MG, Tanese N, Pugh BF, Tjian R. 1990. Functional domains and upstream activation properties of cloned human TATA-binding protein. *Science* 248:1625–30
38. Rice PA. 1997. Making DNA do a U-turn: IHF and related proteins. *Curr. Opin. Struct. Biol.* 7:86–93
39. Rice PA, Yang S, Mizuuchi K, Nash HA. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–1306
40. Richet E, Abcarian P, Nash HA. 1986. The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* 46:1011–21
41. Roeder RG. 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem. Sci.* 16:402–8
42. Saenger W. 1983. *Principles of Nucleic Acid Structure*, ed. CR Cantor. New York: Springer-Verlag. 556 pp.
43. Sawadogo M, Sentenac A. 1990. RNA polymerase B (II) and general transcription factors. *Annu. Rev. Biochem.* 59:711–54
44. Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* 253:1001–7

45. Schumacher MA, Choi KY, Zalkin H, Brennan RG. 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* 266:763–70
46. Seeman NC, Kallenbach NR. 1994. DNA branched junctions. *Annu. Rev. Biophys. Biomol. Struct.* 23:53–86
47. Seeman NC, Rosenberg JM, Rich A. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* 73:804–8
48. Segall AM, Goodman SD, Nash HA. 1994. Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J.* 13:4536–48
49. Steitz TA. 1990. Structural studies of protein-nucleic acid interaction. *Q. Rev. Biophys.* 23:205–80
50. Strauss JK, Maher LJ III. 1994. DNA bending by asymmetric phosphate neutralization. *Science* 266:1829–34
51. Suzuki M, Allen MD, Yagi N, Finch JT. 1996. Analysis of co-crystal structures to identify the stereochemical determinants of the orientation of TBP on the TATA box. *Nucleic Acids Res.* 24:2767–73
52. Suzuki M, Yagi N. 1995. Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.* 23:2083–91
53. Suzuki M, Yagi N, Finch JT. 1996. Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* 379:148–52
54. Tan S, Hunziker Y, Sargent DF, Richmond TJ. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* 381:127–51
55. Thanos D, Maniatis T. 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* 83:1091–1100
56. Tjian R, Maniatis T. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77:5–8
57. Werner MH, Bianchi ME, Gronenborn AM, Clore GM. 1995. NMR spectroscopic analysis of the DNA conformation induced by the human testis determining factor SRY. *Biochemistry* 34:11998–12004
58. Werner MH, Gronenborn AM, Clore GM. 1996. Intercalation, DNA kinking, and the control of transcription. *Science* 271:778–84
59. Werner MH, Huth JR, Gronenborn AM, Clore GM. 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81:705–14
60. Werner MH, Huth JR, Gronenborn AM, Clore GM. 1996. Molecular determinants of mammmalian sex. *Trends Biochem. Sci.* 21:302–8
61. Wolfe SA, Ferentz AE, Grantcharova V, Churchill MEA, Verdine GL. 1995. Modifying the helical structure of DNA by design: recruitment of an architecture-specific protein to an enforced DNA bend. *Chem. Biol.* 2:213–21
62. Wolfe SA, Verdine GL. 1993. Ratcheting torsional stress in duplex DNA. *J. Am. Chem. Soc.* 115:12585–86
63. Yamamoto T, Horikoshi M, Wang J, Hasegawa S, Weil PA, et al. 1992. A bipartite DNA binding domain composed of direct repeats in the TATA box binding factor TFIID. *Proc. Natl. Acad. Sci. USA* 89:2844–48